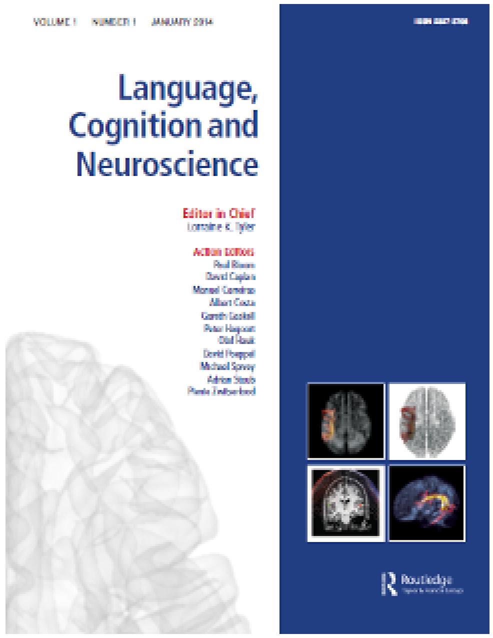
Cross-linguistic gender congruency effects during lexical access in novice L2 learners: evidence from ERPs

Ana Zappa, Daniel Mestre, Jean-Marie Pergandi, Deirdre Bolger, Cheryl Frenck-Mestre



Cross-linguistic gender congruency effects during lexical access in novice L2 learners: evidence from ERPs

Journal:	*Language, Cognition and Neuroscience*
Manuscript ID	PLCP-2020-OP-11063.R3
Manuscript Type:	Original Paper
Date Submitted by the Author:	22-Dec-2021
Complete List of Authors:	Zappa, Ana; Aix-Marseille Universite, Language sciences Mestre, Daniel; Aix-Marseille Universite, Language sciences; Centre National de Recherche Scientifique, PACA, Laboratoire Parole et Langage; CRVM Pergandi, Jean-Marie; Aix-Marseille Universite; CRVM Bolger, Deidre; Institute of Language, Communication and the Brain; Aix-Marseille Universite, Language sciences Frenck-Mestre, Cheryl; Aix-Marseille Universite, Laboratoire Parole et Langage; Centre National de Recherche Scientifique, PACA, Laboratoire Parole et Langage
Keywords:	Gender congruency effect, EEG, L2 learning, Computerized games

Cross-linguistic gender congruency effects during lexical access in novice L2 learners: Evidence from ERPs

Ana Zappa, Daniel Mestre, Jean-Marie Pergandi, Deirdre Bolger & Cheryl Frenck-Mestre

Herein we present electrophysiological evidence of extremely rapid learning of new labels in an L2 (Brazilian Portuguese) for existing concepts, via computerized games. However, the effect was largely constrained by cross-linguistic grammatical gender congruency. We recorded ERPs both prior to exposure with the second language and following a 4-day training session. Results showed rapid changes in cortical activity, associated with learning. Prior to exposure, no modulation of the N400 component was found as a function of the correct match vs. mismatch of audio presentation of words and their associated images. Post-training, a large N400 effect was found for mismatch trials compared to correctly matched audio-visual trials. However, for learners these results were only obtained for trials on which the L2 words shared grammatical gender in the learners L1 (French). For trials on which the L2 words had the opposite gender in French, no N400 mismatch effect was found post-training. In contrast, behavioral results showed that all L2 words were learned equally as well, independent of gender congruency across Portuguese and French. For control participants who were native speakers of Portuguese, a clear N400 effect was found for mismatch compared to match trials, which was independent of gender congruency. The results demonstrate that grammatical gender overlap in the L1 and L2 influences lexical activation during the initial stages of establishing a new L2 lexicon.

In Brazilian Portuguese, a mouse, no matter which biological sex, is grammatically masculine (*o*$_{masc}$ *camundongo*$_{masc}$) whereas a cockroach is grammatically feminine (*a*$_{fem}$ *barata*$_{fem}$). The opposite is true in standard French, with grammatically feminine mice (*la*$_{fem}$ *souris*$_{fem}$) and masculine cockroaches (*le*$_{masc}$ *cafard*$_{masc}$). This arbitrary assignment of grammatical gender is even more apparent for inanimate objects, with opposite gender assignment for trash cans, brooms and chalk across Portuguese and French, despite both languages being derived from Latin. The present study examined how cross-linguistic gender congruency, i.e. the overlap in grammatical gender for nouns across languages, might affect both the acquisition and online processing of a second language (L2) in novice adult learners. Although numerous online studies have provided evidence that speakers of gendered languages are sensitive to gender congruency across languages, during both L2 comprehension and production (cf. Sá-Leite, Fraga & Comesaña, 2019, for a meta analysis) these studies have almost exclusively examined learners who had extensive experience with the L2. We propose a novel approach to this question by starting from the initial stages of exposure to the L2. In addition, we provide both behavioral and electrophysiological measures of performance, which indeed revealed different patterns of the effect of gender congruency.

Numerous psycholinguistic studies have examined the effect of gender congruency (GCE), both within languages in monolinguals and across languages in bilinguals. Monolingual studies have been conducted in the framework of speech production models, which generally assume that grammatical gender is represented independently from other levels of lexical representation, i.e. phonological and semantic, but differ as concerns when and how gender is retrieved (Caramazza, 1997; Cubelli et al. 2005; Foucart et al., 2010; Levelt et al., 1999; Schiller & Caramazza, 2006). Several bilingual studies, discussed below, have also been conducted within this framework. As concerns comprehension, bilingual studies that have examined cross-linguistic GCE have looked at both the interactive nature of bilingual lexical access (Morales et al., 2016; Paolieri et al., 2020) and late bilinguals' ability to use grammatical gender to predict upcoming elements (Hopp & Lemmerth, 2016; Lemmerth & Hopp, 2019). The latter have addressed processing at the lexical level. However, a handful of comprehension studies have also investigated the influence of cross-linguistic gender congruency on syntactic processing (Foucart & Frenck-Mestre, 2011, 2012; Sabourin et al., 2006). We shall address these topics in turn.

Cross-linguistic GCEs have been examined at the lexical level during production, in naming (Bordag, 2004; Bordag & Pechmann, 2007; Costa et al., 2003; Lemhöfer et al., 2008; Morales et al., 2011; Paolieri et al., 2010) and translation (Bordag & Pechmann, 2008; Paolieri et al., 2010; Salamoura & Williams, 2007), as well as in comprehension (Lemhöfer et al., 2008; Paolieri et al., 2020). The pattern of results across studies is both complex and inconsistent. In two independent experiments with German L1-Dutch L2 late bilinguals, Lemhöfer et al. (2008) examined the effects of cognate status and gender congruency on lexical decision times and naming latencies in the L2. They reported robust effects of both factors in both tasks, with no interaction effects. Participants showed faster lexical decision times and naming latencies for L2 Dutch words that shared gender in German and for cognates (cf. Sá-Leite et al., 2019, as well as Costa et al., 2000, for a discussion of cognate effects). In addition, in the naming task, no effect of syntactic structure was found, such that GCE were reported independent of whether participants produced determiner phrases or bare nouns. Similar results were reported by Bordag & Pechman (2007) for relatively inexperienced Czech L1-German L2 late learners, who found GCE in 2 experiments independent of whether participants named bare nouns or nouns preceded by gender-marked adjectives in the L2. In like fashion, Paolieri and colleagues reported faster naming latencies in the L2 (Spanish) for line drawings that shared the same gender in the bilingual participants' L1 (Italian) (eg "falda" and "gonna" *skirt*) compared to those that had opposite gender (e.g., "mesa" and "tavolo" *table*) and this was true for both bare nouns and determiner noun phrases (Paolieri et al., 2010). Paolieri et al. (2010) replicated their results for bilingual participants across 2 experiments; however, in their second experiment both the effect of gender congruency and the critical interaction with Group (bilingual vs. monolingual) were only reliable by participants, suggesting that the GCE was not restricted to the bilingual group and may have been partially driven by characteristics of the items unrelated to gender congruency. Klassen (2016) examined GCE in naming across Spanish and German, which differ concerning the number of gender categories (2 vs. 3), the phonetic regularity of gender marking (overt and highly consistent in Spanish vs. not phonetically salient in German) and complexity (case interacts with gender in German but not Spanish). Results showed facilitated processing for gender congruent nouns compared to incongruent nouns, for both bare nouns and determiner phrases in the group of L1 Spanish-L2 German learners but not L1 German controls. Yet, this was true for nouns that were either masculine or feminine in German but not for neuter gender, which, although indeed incongruent with either gender in Spanish led to faster naming latencies. Hence, the GCE also appears to be constrained by the overlap of gender categories across languages. All of these studies argue in favor of an interactive activation model of processing (Dell, 1986) in which both the L1 gender and phonological form of lexical entries influence L2 processing, and according to which grammatical gender is not stored as an independent feature at the lemma level (i.e. Levelt et al., 1999).

In contrast, Costa and colleagues failed to find GCE across 5 independent experiments in which participants produced NPs in their L2, even when with gender-marked determiners (Costa et al., 2003). This was true independent of whether the bilinguals' two languages had similar gender systems. Costa and colleagues (2003) argued that while semantic representations are shared across languages and commonly activated by lexical entries of either language, the specific grammatical features of a lexical entry, such as its gender, are inherent properties of that entry. Hence, these features would not be shared across languages. Costa et al. (2003) noted nonetheless that they tested highly proficient bilinguals and suggested that less proficient L2 speakers might show greater interaction between their two gender systems, as was indeed reported by subsequent studies (Bordag & Pechmann, 2007; Lemhöfer et al., 2008; Paolieri et al., 2010). In addition, as highlighted by Sá-Leite et al. (2009), other mitigating factors that were not considered by Costa et al. (2003) may have played a role and could explain why this study seems to be the odd man out as concerns finding GCE.

In translation tasks, the results are also mixed. At one extreme, Bordag & Pechmann (2008) reported no cross-linguistic GCE across three translation tasks with Czech-German participants who translated either bare nouns or adjective-noun phrases into the L2. It is of particular interest that the absence of a GCE was reported in translation for the same materials and participant population that produced a robust GCE in production (Bordag & Pechmann, 2007). Bordag and Pechmann (2008) account for the discrepancy in results across tasks as concerns GCE in terms of time course. During picture naming, activation would spread from the concept to the L1 and L2 lemmas in parallel, hence leading to the simultaneous activation of L1 and L2 gender nodes and competition when the two do not match. In translation, the time course would be shifted, such that L1 word forms would activate their lemmas which in turn would activate the L2 lemma and word form. As such, the L2 gender node would only be activated subsequent to the L1 gender node and no direct competition would arise. Salamoura & Williams (2007) reported a different pattern of results, whereby proficient Greek-German bilinguals showed cross-linguistic GCE when they translated gender-marked adjectives along with the noun, but not for bare nouns. The authors argued that gender retrieval occurs only when gender concord must be computed, i.e. within the determiner phrase, in line with certain monolingual models of production (Caramazza, 1997, but see Cubelli et al., 2005). Last, Paolieri et al. (2010) reported GCE during a translation task with advanced Italian-Spanish bilinguals, independent of whether participants produced the bare noun or a determiner phrase. Based on this pattern of results, the authors proposed that lexical selection necessarily entails the activation of gender, in parallel. Lexical items that share gender across languages would enjoy a higher level of activation, hence facilitating both naming and translation. The three competing sets of results and theoretical accounts leave ample room for discussion.

The effect of gender congruency across languages has also been examined during online comprehension. Several studies have approached this topic in the framework of whether bilinguals can use grammatical gender to predict upcoming elements in their L2 (Hopp & Lemmerth, 2016; Lemmerth & Hopp, 2019; Morales et al., 2016). The results from two visual world paradigm studies with Russian-German bilinguals, which used the same design and materials, failed to produce statistically conclusive evidence that gender congruency plays a significant role in the ability of either adults (Hopp & Lemmerth, 2016) or children (Lemmerth & Hopp, 2019) to process gender online in the L2. Another visual world paradigm study, conducted with proficient Italian-Spanish bilingual adults, showed interference from a distracter image when its gender was incongruent with the L1 equivalent, suggesting gender-induced competition (Morales et al., 2016). However, the effects were not significant until after the onset of the target noun, suggesting that co-activation of gender across languages during comprehension may not occur until a certain amount of information has been processed.

Event-related potentials (ERPs) have also been used to measure the effects of cross-linguistic gender congruency on language processing. The most common of these are the N400, the P600 and the LAN, generally elicited by semantic and syntactic violations respectively. The N400, a negative deflection in the waveform with a central-parietal distribution usually observed between 300-500 msec after stimulus onset, is generally thought to reflect semantic integration such that increased N400 amplitude is attributed to processing difficulty that results from attempting to integrate a new element within an existing semantic context (Holcomb, 1993; Kutas & Federmeier, 2011; Kutas & Hilliyard, 1980). The N400 component has also been associated with retrieval/access, with increased N400 amplitude reflecting the effort of retrieving semantic/conceptual information from long-term memory. In this case a reduced N400 is interpreted as signaling facilitated access to lexical information (Delogu et al., 2019). Recently, Bornkessel-Schlesewsky & Schlesewsky (2019) re-examined N400 results in a predictive coding perspective. Accordingly, the N400 (along with other language-related negativities) could reflect precision-weighed prediction errors rather than linguistic processing per se. Another ERP component commonly used in language studies is the P600, which is a positive-deflection in the wave form with a centro-parietal distribution thought to show difficulty in syntactic integration (Kaan et al., 2000; Meltzer & Braun, 2013), as well as the cost of "revising, repairing or reanalyzing an existing (morpho-) syntactic structure" (Delogu et al., 2019, p.2). The P600 is sometimes associated with the Left Anterior Negativity (LAN) (for a debate on the significance of the LAN, cf. Fraga et al., 2021; Molinaro et al., 2015; Tanner, 2015). The P600/LAN has also been associated with conflict/monitoring resolution (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kim & Osterhout, 2005) as well as semantic integration (Brouwer et al., 2017). Finally, the nogo N200 response occurs roughly 200 msec post-stimulus when there is conflict and/or inhibition as concerns processing and the participant's response (Enriquez-Guppert et al., 2010). While not specific to language processing, the N200 can be used to measure response conflict during language processing such as in the study described below.

Rodriguez-Fornells & Münte (2016) recorded ERPs in a Go/Nogo paradigm to probe the effect of grammatical gender congruency (and language switching) across German and Spanish in fluent bilinguals. Compared to monolingual controls, bilinguals showed greater negativity (N200) for incongruent compared to congruent gender trials. While this result may indicate the automatic activation of gender across languages, there are several caveats. First, the task explicitly required participants to retrieve grammatical gender and both languages were actively recruited. Second, participants could potentially predict the incongruent gender trials based on the structure of the experiment. Last, gender incongruence elicited a late ERP component (P600/LPC). Hence, these results do not provide clear evidence for the automatic, early retrieval of gender. Paolieri et al. (2020) also recorded ERPs to examine the effect of gender congruency during the processing of translation equivalents in Spanish and Catalan. Only bare nouns were presented. A small but reliable increase in N400 amplitude was found at central-midline sites for translation pairs that did not share grammatical gender across languages in comparison to those that did, along with increased response times for incongruent pairs, leading the authors to claim that gender is automatically activated during lexical retrieval and elicits competition when different genders are activated across languages. As with Rodriguez-Fornells & Münte (2016), however, participants were actively processing both languages; a stronger demonstration would have consisted in comparing N400 amplitude for bare nouns in the L2 alone as a function of gender congruency. Moreover, no L1 control group was included such that it is not possible to determine whether the effect was driven solely by gender congruency or perhaps by extraneous factors specific to the gender-incongruent pairs. Nevertheless, in a semantic categorization task conducted exclusively in English with Spanish-English late bilinguals and monolingual controls, Boutonnet et al. (2012) reported a late negative ERP component (starting at roughly 400 msec) for trials that did not share gender in Spanish (the L1) with the two preceding items. This effect was specific to the bilingual group. No effects were found for the behavioral measures, in contrast to Paolieri et al. (2020), although the absence of an effect for behavioral measures in the presence of electrophysiological evidence is rather common. The authors argued for the automatic activation of L1 gender in speakers of gendered languages, even when processing exclusively the L2 and in a non-gendered language such as English.

Gender congruency effects have also been examined at the syntactic level during sentence processing, using ERPs. Several monolingual studies have shown that gender agreement violations in sentential context systematically elicit the P600 component (Alencar de Resende et al., 2019; Barber & Carreiras, 2005; Beatty-Martinez et al., 2021; Hagoort, 2003; Frenck-Mestre, 2005; Fraga et al., 2021; Gunter et al., 2000; Popov & Bastiaansen, 2018; Popov et al., 2020; Wicha et al., 2004) and can also produce a LAN (Alencar de Resende et al., 2019; Barber & Carreiras, 2005; Beatty-Martinez et al., 2021; Fraga et al., 2021; Gunter et al., 2000; Popov et al., 2020). The electrophysiological signature of gender concord during sentence processing is nonetheless modulated by various factors, both linguistic (Alencar de Resende et al., 2019; Beatty-Martinez et al., 2021) and task-related (Schacht et al., 2014). In Brazilian Portuguese, Alencar de Resende and colleagues (2019) compared ERP signatures elicited by gender concord violations within the determiner phrase, either between the determiner and noun or the adjective and noun, for nouns with either regular or irregular gender assignment in Brazilian Portuguese. Results showed a biphasic LAN/P600 in response to concord violations, independent of the regularity of gender assignment. However, the amplitude of the P600 evoked by concord violations was greater for regular nouns. The authors suggested that both regular and irregular forms are stored and accessed in similar fashion, i.e. via a single lexical route, but that repair processes are facilitated for regular forms. Beatty-Martinez et al. (2021) reported differences as a function of gender category in Spanish, whereby determiner-noun violations in sentence contexts elicited a biphasic LAN/P600 response for masculine nouns in contrast to the same violations for feminine nouns, which did not elicit a LAN and showed a wider P600 distribution in comparison to masculine nouns. This difference in the electrophysiological response was interpreted as being linked to systematic differences in the representation of the two gender categories.

The P600 is also elicited during L2 sentential processing of gender concord violations for L2 learners whose L1 has grammatical gender (Foucart & Frenck-Mestre, 2011; Sabourin et al., 2006) but also for those who do not (Foucart & Frenck-Mestre, 2012; Dowens et al., 2011; Morgan-Short et al., 2010; Tokowicz & MacWhinney, 2005). Moreover, P600 amplitude for gender concord violations in an L2 is contingent on proficiency and age of acquisition (Nichols & Joanisse, 2016). ERP studies that focused on cross-language gender congruency have provided evidence that the overlap of both syntactic rules, as concerns gender concord (Foucart & Frenck-Mestre, 2011;

Sabourin & Stowe, 2008), and lexical gender across languages (Foucart & Frenck-Mestre, 2011) affect whether gender concord violations in the L2 elicit an electrophysiological response, the type of response (Foucart & Frenck-Mestre, 2012) and its magnitude.

The body of studies cited above has examined L2 gender processing and gender congruency in participants who had several years of experience with and exposure to the L2. Various authors have used learning paradigms with either an artificial language or miniature versions of natural languages to explore how different factors affect gender acquisition after short training periods. Arnon and Ramscar (2012) used an artificial language to test whether acquiring the gender and new lexical labels of known concepts was affected by the sequence of explicit training. Participants who first learned novel words preceded by their gender-marked article within an auditory sentential context, followed by paired associate learning with bare nouns had better learning outcomes, both for gender assignment and noun labels, than those who learned in the opposite order. The authors argued that learning new lexical labels for concepts first via bare nouns blocked the later acquisition of gender assignment in sentential context due to the redundancy of gender in relation to meaning. Brooks and Kempe (2013) examined the implicit learning of nominal gender agreement and case marking in a subset of Russian following a 6-day training session. Results showed that while learners relied on metalinguistic knowledge to acquire case marking, for gender concord they relied on the consistent and transparent morphological cues (feminine being systematically indicated by a final vowel on the noun and agreeing adjective, and masculine by a final consonant) and knowledge of nominal morphology in another known L2. Indeed, the best predictor of acquiring Russian gender agreement was whether the learners had already acquired a Latin language with the same rule for feminine gender. Morgan-

Short and colleagues (2010) compared the processing of gender concord in early learning stages, after implicit versus explicit training in an artificial language. During early stages of acquisition, the ERP signature to gender concord violations differed according to the type of training. Notwithstanding, both implicit and explicit learning groups ultimately attained similar levels of proficiency and exhibited similar patterns of cortical response to gender concord violations at the final stage of acquisition.

To our knowledge, no studies have measured the effect of cross-language gender congruency during the early stages of L2 lexical acquisition in a natural language. The current study aimed to fill this void by examining how cross-linguistic gender congruency might influence processing in an L2 from the very initial stages of learning. We examined French L1 speakers' acquisition of Brazilian Portuguese via interactive computer games, in which L2 Portuguese was presented aurally in full sentences and segmented format, in which grammatical gender within determiner phrases was taught implicitly. Both French and Portuguese have two classes of grammatical gender (masculine and feminine) and require gender concord within the determiner phrase. Whereas French uses the singular definite article le[lə]$_{masc}$ to mark the masculine gender and Portuguese uses o$_{masc}$ (realized as /o/ or /u/), in both languages the singular feminine definite article carries the final phoneme [a] (French: la$_{fem}$ [la], Portuguese a$_{fem}$ [a]). In addition, in Portuguese, the vowel of the definite determiner is generally consistent with the final vowel of the noun (e.g.,"a faca" "the knife" and "o garfo" the fork). It is therefore probable that, even without formal instruction concerning the gender of the Portuguese nouns or the determiner system, French native speakers are able to extract this information from the phonological word forms (Brooks & Kempe, 2013; Denhovska & Seratrice, 2017). Indeed, following the seminal work by Karmilloff-Smith (1979), various experimental studies have shown that native French speakers are sensitive to the regularities present in noun endings and reliably use them as early as age 3 to process and/or assign grammatical gender (cf. Seigneuric et al., 2007 for a discussion and further work with children, and Pérez-Pereira (1991) for similar work in Spanish). In Brazilian Portuguese, incongruent grammatical gender marking within the determiner phrase between adjacent elements (notably the determiner and noun) affects children's ability to process gender assignment as young as age 2 (Corréa & Name, 2003), in support of the hypothesis that young children are sensitive to grammatical information conveyed by the determiner and process gender concord in the DP (cf. Corrêa et al., 2011 for further work). Thus, native speakers of gendered languages show rapid acquisition of the phonological features associated with grammatical gender categories, when present, and the syntactic reflexes of elements that are controlled by the noun. Our French learners undoubtedly exploited these capacities when acquiring the L2 Brazilian Portuguese vocabulary.

We created 4 computer games to teach French native speakers a small lexicon in Brazilian Portuguese. All auditory materials were presented exclusively in Brazilian Portuguese. The games involved both full sentences and individual lexical items, which comprised noun phrases (definite determiner and noun) and verbs. We manipulated gender congruency such that the nouns were either gender congruent or incongruent across the learners' L1 and L2 (cf. Table 1). No instruction was provided concerning grammatical gender, however; nouns were always preceded by the singular definite determiner, i.e. the overtly marked vowel (e.g., a_{fem} $faca_{fem}$ [the knife], o_{masc} $garfo_{masc}$ [the fork]). Participants with no prior knowledge of (Brazilian) Portuguese took part in a 3-day training program, during which they learned a vocabulary of 36 words overall, divided into 12 sentences comprised of 3 verbs and 4 nouns each day via the games. ERPs were recorded both pre- and post-training in a match/mismatch paradigm in which auditory nouns were paired with visual images that either depicted the noun (match) or another learned noun (mismatch). Thus, we were able to follow progression from zero knowledge to the recognition of newly learned L2 phonological word forms.

We hypothesized that learners should be able to fully acquire the L2 vocabulary. Performance was measured both by their accuracy scores pre- and post-training and, crucially, the change in electrophysiological response pre to post-training. Concerning the latter, we expected variation in the N400 component, whereby prior to training match and mismatch trials should not differ in the N400 response, but post-training mismatch trials should evoke an increased N400 response compared to match trials due to difficulties in lexical processing. In relation to the congruency of grammatical gender across the L1 and L2, we hypothesized that it should not affect the ERP response pre-training. In contrast, at the post-training session, it should affect the size of the N400 effect if indeed grammatical gender is automatically activated for speakers of gendered languages (Boutonnet et al., 2012; Dahan et al., 2000; Lew-Williams & Fernald, 2007) and if inhibition results from gender inconsistency across languages (Morales et al., 2016; Rodriguez-Fornells & Münte, 2016).

	L2fem	L2masc
L1fem	a saia (la jupe)	o cachimbo (la pipe)
L1masc	a vassoura (le balai)	o casaco (le manteau)

Table 1. Examples of gender congruency/incrongruency across Brazilian portuguese and French

Methods

Participants. Eighteen right-handed French native speakers (10 women), enrolled as undergraduate students at Aix-Marseille Université, aged 20 to 26 years old (M = 22.8 years, *SD* = 2.4), who had no knowledge of Brazilian Portuguese, and 18 right-handed native speakers of Brazilian Portuguese (16 women) aged 22 to 28 (M = 25.3, SD = 3.4) who were enrolled at AMU in a one year abroad program were recruited for the study. One French participant's data was excluded due to displaying knowledge of the L2 vocabulary (an N400 effect for mismatched pairs) prior to training. All French participants had learned English as a second language throughout secondary school; eight had also learned Spanish as a third language while seven had learned German. None considered themselves fluent in Spanish. The Brazilian participants had all learned English in secondary school; 12 considered themselves novice in French and 6 intermediate. Participants had no history of neurological insult and received monetary compensation in exchange for their participation. All participants gave their written informed consent prior to the experiment and were debriefed about its purpose at its end. The study was approved by the local university ethics committee.

Materials. Thirty-six concrete nouns and 9 transitive verbs were presented orally in Brazilian Portuguese (BP), in sentences and in isolation, across 3 training sessions, with 3 verbs and 12 nouns learned in each session. All materials are presented in the appendix. The items were selected based on their ease of imageability, cognate status with the learners' L1 (French) as well as other Latin languages commonly learned in France (Italian, Spanish) and English, and congruency of grammatical gender across Portuguese and French. Half of the nouns had the same gender across French and Brazilian Portuguese ([a_f $panela_f$] / *la*$_f$ *casserole*$_f$ *"the pot"*) and the other half had the opposite gender ([a_f $faca_f$] / *le*$_m$ *couteau*$_m$ *"the knife"*). Of the 36 nouns, 32 were phonetically marked for gender (16 ending with /a/ and 16 ending with /o/) while 4 were phonetically opaque (2 feminine and 2 masculine nouns). As an added precaution, the overlap in gender between Portuguese and Spanish was also checked for the set of 36 nouns. For the 18 nouns that had the same gender in French and Portuguese, 14 also matched between Portuguese and Spanish and 14 between French and Spanish. For the 18 nouns that had opposite gender across French and Portuguese, 7 had opposite gender between Portuguese and Spanish and 11 had opposite gender between French and Spanish. Nouns were systematically preceded by the definite determiner (/a/$_f$ or /o/$_m$). Nouns that shared gender (SG) and those that had opposite gender (OG) across French and Portuguese were equated across numerous lexical variables: printed mean frequency per million in French (M = 19.71, sd = 21.19) vs. (M = 18.04, sd = 16.05) for SG and OG, respectively (New, Pallier Brysbaert & Ferrand, 2004), mean number of letters SG (M = 6.17, sd =1.70) vs. OG (M = 5.53, sd = 1.62), mean number of phonemes SG (M = 5.50, sd = 1.62) vs. OG (M = 5.47, sd = 1.51) and mean number of syllables SG (M = 2.58, sd = 0.79) vs. OG (M = 2.47, sd = 0.80), grammatical gender SG (9fem/6masc, tested in the Match/mismatch task) vs. OG (8fem/7masc, tested in the Match/mismatch task) and Levenshtein distance SG (M = 5.5, sd = 1.4) vs. OG (M = 6.0, sd = 1.62).

Each of the 9 verbs was paired with 4 different nouns to create 36 declarative sentences in canonical SVO order preceded by a lead in phrase e.g., Esfregar / Scrub [Ele está esfregando a janela / a lareira com a escova / o trapo] / He is scrubbing the window/fireplace with the brush/rag)). Three additional partially transparent verbs (e.g., “pintar” “peindre” paint) and 12 additional partially transparent nouns (e.g., “esponja”, “éponge”, sponge) were selected to familiarize participants with the games and for EEG training. All auditory materials were recorded by a native Brazilian female speaker at 48 kHz (32-bit float) in a professional sound booth, in a single session. The onset of each word within auditory sentences was detected automatically using SPPAS (Bigi, 2015) and manually verified using PRAAT (Boersma & Weenink, 2018). The materials were spliced into individual syntactic units (pronoun + copula, determiner + noun, lexical verb) and individual sentences using Audacity 2.2.1 software.

A subset of the materials was selected for pre- and post-training in a Match/mismatch task (see appendix). Thirty auditory nouns were presented in a 2x2 factorial design defined by the congruency of the auditory noun and line drawing (match vs. mismatch) and gender congruency across languages (same vs. opposite gender). Auditory nouns were paired with 30 line drawings selected from the Snodgrass & Vanderwart (1980) and Alario and Ferrand (1999) standardized image databases (with the exception of two images which were taken from line-drawing internet databases). These images were different from those used for the computerized games (described below), which were selected from internet databases. Each auditory noun was presented twice, once paired with the correct line drawing (match) and once with a line drawing that corresponded to another (to be) learned noun (mismatch). For each Condition (March vs. Mismatch) half of the trials had the same grammatical gender across Portuguese and French and half had opposite gender: Gender (Same vs. Opposite). Hence, a full factorial design was used: Gender (Same vs. Opposite) x Condition (Match vs. Mismatch). For Match trials, where the presented image matched the auditory word, only the gender of the auditory word was in play, which was either the same in Portuguese and French (e.g., *a janlea / la fenêtre* [the window]) or opposite across languages (e.g., *o garfo / la fourchette* [the fork]). For mismatch trials, where the image *did not* correspond to the auditory word, we controlled for the coherency of gender across the distractor image and auditory target word in Portuguese. For example, the auditory word *o cachimbo* (*la pipe* [the pipe]) was paired with the distractor image of a masculine Portuguese word *o lixo* (*la poubelle* [the trashcan]). This prevented participants from noting a mismatch based on the determiner alone if they accessed the name of the distractor image prior to the auditory target word. Three pseudo-randomized lists were created for pre-training EEG testing including 30 "match" pairs and 30 "mismatch" pairs, with 15 same gender and 15 opposite gender pairs for each. Three other pseudo-randomized lists were created for post-training EEG testing, which included the same 30 Match and 30 Mismatch pairs and an additional 30 Semantically related and 30 Semantically unrelated pairs (data reported elsewhere). Participants saw different lists at pre- and post-training testing, with complete counter-balancing of lists across participants such that all lists were seen at both pre- and at post-training sessions.

Procedure

Games. Four computerized games were created in collaboration with the Mediterranean Virtual Reality Center (CRVM). All 4 games involved the auditory presentation via headphones of materials in Brazilian Portuguese accompanied by either static line drawings or animated GIF on a flat screen. Participants' responses and playing behavior (mouse clicks, timing) were recorded throughout each game and feedback was provided. Trials that were not completed successfully were repeated at the end of each game. All participants played the 4 games in the same order and were required to successfully complete a given game prior to engaging in the next. This ensured that all participants had acquired the vocabulary presented throughout the games on each day. In all 4 games, participants initiated a trial by clicking on an audio button image to hear an auditory stimulus. In the first, "exposure" game, participants clicked to hear the 12 sentences, one at a time, which were presented simultaneously with an animated GIF of the action and accompanying objects. In the second "segmentation" game, participants clicked to hear a sentence, which was accompanied by the visual presentation of 5 blank squares at the bottom and 3 in the center of the computer screen. Participants clicked on any of the 5 bottom squares to display a static image and hear the audio file corresponding to it (4 nouns and 1 verb were depicted). Participants had to recreate the auditory sentence by clicking and dragging the 3 correct syntactic elements in order (verb, NP1, NP2) to the center of the screen. Upon correct completion, an animated GIF played along with the auditory sentence. In the third, "verb identity" game, participants clicked to hear a verb and saw three different animated GIFs in the center of the screen, the task being to select the animation that corresponded to the audio file. In the final "memory" game, participants clicked to hear a sentence, then had to find the pair of cards, among 8 presented face down on the screen, that matched the auditory sentence by clicking on the cards individually (a card reverted to blank when another card was selected). The vocabulary depicted across the 8 cards involved a single verb and 4 nouns. When the correct pair was selected, the auditory sentence was replayed along with an animated GIF. Across games 2, 3 and 4, participants were allowed to click on a given item (audio button or card) a maximum of 3 times and time-out was 30 seconds on any given trial. They were encouraged to repeat the materials out loud while playing. Participants were not permitted to take notes during training sessions and were asked not to review what they had learned between sessions. All participants successfully completed all games on each of the three training days, as revealed by their recorded scores.

Training. Participants learned how to play the 4 computer games, prior to actual training, using a miniature auditory vocabulary. This initiation to the games took place directly after the first EEG session. During the initiation and subsequent training, participants were comfortably seated in a sound-attenuated room where they played the games on a 15-inch-screen laptop computer while wearing headphones. Training occurred over 4 consecutive days. On each of the first three days, participants were exposed to 3 verbs and 12 nouns, comprised in 12 auditory sentences, via the 4 games, with the vocabulary repeated across the games. Each session lasted roughly 25 minutes, with each game taking 5-10 minutes. The fourth day consisted of a 40-minute review day, where participants played all the games with the entire new lexicon (9 verbs and 36 nouns).

EEG Match-Mismatch task. A trial began with the presentation of a centered fixation cross for 250 msec that was replaced by a centered black and white line drawing for 1 second followed by an auditory word presented over speakers. At the offset of the auditory word, a visual "yes/no" prompt was presented and participants were requested to judge whether the auditory word matched the visual image or not on a button box. A visual blink prompt was presented for 2s following the response. During the pre-training session, 2 blocks of 30 trials, with 15 "match" and 15 "mismatch" pairs in each, were presented, preceded by three warm-up pairs. During the post-training session, an additional 30 trials were presented per session, comprising 15 semantically related and 15 unrelated pairs (data reported elsewhere), for a total of 60 trials per session. Short rest periods were allowed between blocks. Participants were asked to remain still and to blink at the prompt. Behavioral responses to the questions were recorded. The session lasted roughly 30 minutes.

ERP Data acquisition and analysis. Electroencephalographic (EEG) activity was recorded continuously from 64 scalp locations over frontal, temporal, central, posterior temporal, parietal and occipital areas of the left and right hemispheres and midline.

Individual electrodes were adjusted to a stable offset lower than 20μV. EEG data were sampled online at 512 Hz. Blinks and horizontal eye-movements were monitored by means of electrodes placed beneath the left eye and at the outer canthus of the right eye. Electrodes were placed on the left and right mastoids for referencing offline. Periods spanning from -100 pre- stimulus onset to 1100 msec post stimulus onset were used post-recording for analyses. A low pass digital filter of 30 Hz was applied post-recording. Trials contaminated by ocular-motor or muscular artifacts were excluded using automated routines that were manually checked. The percentage of trails retained for analyses was 88% for the same gender condition and 87% for the opposite gender condition.

Results.

To determine the pattern of data, we ran a series of comparisons for both behavioral and electrophysiological measures. One the one hand, we compared the data for the L2 learner group across sessions. In addition, we compared the L2 learners' data post-training to that of the Brazilian control group. When indicated, we ran subsequent independent models (in each session, for the L2 learners and for each group).

Behavioral measures.

Game-playing accuracy

Errors took the shape of time-outs. Accuracy was at ceiling (over 96% across all games) for all participants. The low level of variability during game play in the review session did not warrant further analyses (cf. Table 2a).

Verb game	Segmentation game	Memory game
22/648 (3.4%)	3/864 trials (0.3%)	3/864 trials (0.3%)

Table 2a. *Total number and percentage of time-outs across the different games for all L2 participants*

Accuracy in the EEG match-mismatch task.

The percentages of correct responses are presented in Table 2b and d prime scores are presented in Figure 1 for the L2 learners at the pre- and post-training session, and for the Brazilian control group. The data were modeled using linear mixed effect regressions, with the *LmerTest* package (Kuznetsova & Christensen, 2017) implemented in R (R Core Team, 2017). For both measures, we first modeled the data for L2 learners, pre- and post-training. The model for the percentage of correct responses included the sum coded fixed factors Training session (Pre vs. Post), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions, with random intercepts for Participant and Item and a random slope for Condition:Gender. The model for d prime included the sum coded fixed factors Training session (Pre- vs. Post), Gender (Same vs. Opposite) and their interaction, with random intercept for Participant. We subsequently compared L2 learners to Brazilian controls at the post training session. For accuracy, the model included the sum coded fixed factors Group (Control vs. L2), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions, with random intercepts for Participant and Item. The same model was applied for d prime, without the factor Condition.

Accuracy.

For L2 learners, the first model, comparing pre- and post-training sessions revealed an effect of Session ($\beta = 1.35$, se = 0.07, z = 18.85, $p < .001$), the interaction between Gender and Condition ($\beta = 0.19$, se = 0.08, z = -2.25, $p < .001$) and the interaction between Gender, Condition and Session ($\beta = 0.14$ se = 0.07, z = 2.02, p

< .05). Subsequently, models were run on the L2 data for the pre- and post-training session independently. Pre-training, the treatment coded model, revealed an interaction between Gender and Condition (β = -1.32 se = 0.28, z = -4.74, $p < .001$), and a simple effect of Condition for Opposite gender trials (β = 0.51, se = 0.19, z = 2.74, $p < .01$). The re-leveled treatment coded model showed a simple effect of Condition for Same gender trials (β = -0.81 se = 0.21, z = -3.92, $p < .001$). Pre-training, L2 participants showed a bias to respond positively for Same gender trials and negatively for Opposite gender trials, although accuracy remained at chance levels. Post-training, the sum coded model revealed only an effect of the Intercept for L2 learners (β = 2.78, se = 0.26, z = 10.68, $p < .001$), due to accuracy being higher than chance. No other effects were significant.

Last, the model comparing L2 learners' performance at the post-training session to the native Brazilian control group failed to converge due to the extremely low variability and high level of accuracy.

	Match	**Mismatch**
L2: Pre-training		
Same Gender	57% (49)	38% (48)
Opposite Gender	36% (48)	47% (50)
L2: Post-training		
Same Gender	92% (27)	92% (29)
Opposite Gender	92% (27)	90% (27)
Brazilian control group		
Same Gender	95% (21)	100% (0)
Opposite Gender	98% (26)	100% (0)

Table 2b. *Percentage of correct answers (standard deviations in parentheses) in the Match-Mismatch task for L2 learners pre- and post- learning, and for Brazilian controls, as a function of Gender (same vs. opposite) and Condition (Match vs. Mismatch)*

D-prime.

The model comparing L2 learners' data pre- and post-training revealed only an effect of Session (β = 1.76, se = 0.105, t = 16.90, $p < .001$). Participants' d prime scores (i.e. their ability to correctly discriminate between Match and Mismatch trials) increased significantly from pre- to post-training and this did not depend on the gender congruency of trials.

The model comparing d prime scores for L2 learners post-training and the native Brazilian control group revealed an effect of Group ($\beta = 0.59$, se = 0.102, $t = 5.80$, $p < .001$), due to Brazilian controls showing higher d prime scores. No other effects were significant.

Figure 1 about here

Figure 1. *D prime scores for Brazilian controls and for L2 learners at pre- and post-training sessions*

ERP analysis

The ERP data were modeled using linear mixed effect regressions, with the *LmerTest* package (Kuznetsova & Christensen, 2017) implemented in R (R Core Team, 2017) for the mean voltage amplitudes in the N400 time window, calculated 300-600 msec post onset of the auditory noun for correct trials. This window was based on prior literature and confirmed by permutation tests conducted across the entire epoch (see below). Data were trimmed in R to remove outliers (1% of the data were excluded). Models were performed independently over midline sites (Fz, FCz Cz, CPz, Pz), frontal-central sites (FC1, FC3, FC5, FC2, FC4, FC6, C1, C3, C5, C2, C4, C6) and centro-parietal sites (CP1, CP3, CP5, CP2, CP4, CP6, P1, P3, P5, P2, P4, P6). Below we report the results from the maximal random-effects structure (Barr, et al., 2013).

In addition to the LMER models, to test the hypothesis of a significant difference in the ERP amplitude between conditions, a two-tailed permutation test (1000 random partitions) was carried out over the 1-second post-stimulus time window for all electrodes entered into the models. Statistically significant differences were taken into consideration only if they persisted for 10 msec or more, which corresponds to an interval of 5 samples, given a sampling rate of 512Hz. The results of these tests are visible in Figures 2a through 4b for 9 central electrodes commonly associated with the N400 effect and voltage maps that included a larger array of electrodes.

Pre vs. Post-training: L2 learners

To examine the effect of training, we ran a first model, summarized in Table 3a, which included the sum-coded fixed factors Session (Pre vs. Post), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions, with random intercepts for Participant and Item. Condition included a random slope for Participant and for Item. The model revealed a three-way interaction of Condition:Gender:Session at all electrode sites. The data were modeled independently thereafter for each training session. Pre-training, no effects were found for any factor at any ROI (cf. Table 3b). Figures 2a and 2b show the mean Match-Mismatch ERP response for Same and Opposite gender conditions, respectively, for illustrative purposes.

Figure 2a about here

Figure 2a. *Mean voltage ERPs (and SD) for 9 central electrodes in the pre-training session for L2 learners as a function of Condition (Match vs. Mismatch) for nouns with same gender across languages. Permutation tests revealed no significant differences.*

Figure 2b about here

Figure 2b. *Mean voltage ERPs (and SD) for 9 central electrodes in the pre-training session for L2 learners as a function of Condition (Match vs. Mismatch) for nouns with opposite gender across languages. Permutation tests revealed no significant differences.*

Post-training, sum coded models performed independently at all 3 ROI revealed a significant interaction of Condition x Gender at all sites (cf. Table 3c). Models of simple effects (cf. Tables 3d and 3e) revealed a significant effect of Condition for same gender trials at all ROI (midline: $\beta = -3.40$, se = 0.770, $t = -4.42$, $p < .001$; frontal-central: $\beta = -2.60$, se = 0.682, $t = -3.82$, $p < .002$; central-parietal: $\beta = -2.91$, se = 0.577, $t = -5.05$, $p < .001$) but no effect of Condition for opposite gender trials (midline: $\beta = -0.27$, se = 0.757, $t = -0.36$ $p<0.73$; frontal-central: $\beta = -0.52$, se = 0.644, $t = -0.81$,$p < .43$; central-parietal: $\beta = -0.59$, se = 0.613, $t = -0.97$, $p < .35$). These effects are shown in Figures 3a and 3b.

Figures 3a about here

Figure 3a. *Mean voltage ERPs (and SD) for 9 central electrodes in the post-training session for L2 learners as a function of Condition (Match vs. Mismatch) for nouns with same gender across languages. Permutation tests are shown in red (p=.05) and green (p<.05) across the entire epoch.*

Figures 3b about here

Figure 3b. *Mean voltage ERPs (and SD) in the post-training session for L2 learners for 9 central electrodes as a function of Condition (Match vs. Mismatch) for nouns with opposite gender across languages. Permutation tests revealed no significant effects.*

In addition, for Match trials there was an effect of Gender at all three ROI (cf. Table 3e) (midline: $\beta = 2.08$, se = 0.422, t = 4.92, $p < .001$; frontal-central: $\beta = 1.69$, se = 0.584, t = 2.89, $p < .01$; central-parietal: $\beta = 1.27$, se = 0.257, t = 4.96, $p < .001$), while for Mismatch trials the effect of Gender was only present at midline (midline: $\beta = -0.91$, se = 0.434, t = -2.09, $p < .04$; frontal-central: $\beta = -0.38$, se = 0.805, t = -0.48, $p<0.64$; central-parietal: $\beta = 1.06$, se = 0.89, t = -1.19, $p < .0.25$). These effects are depicted in Figures 3c and 3d.

Figure 3c about here

Figure 3c. *Mean voltage ERPs (and SD) for 9 central electrodes in the post-training session for L2 learners as a function of Gender across languages (same vs. opposite) for Match trials. Permutation tests are shown in red (p=.05) and green (p<.05) across the entire epoch.*

Figure 3d about here

Figure 3d. *Mean voltage ERPs (and SD) for 9 central electrodes in the post-training session for L2 learners as a function of Gender across languages (same vs. opposite) for Mismatch trials. Permutation tests revealed no significant effects.*

Post-training: Native controls vs. L2

The first model, summarized in Table 4a, included the sum-coded factors Group (Native vs. L2), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions. Condition included a random slope for Participant and for Item. The model revealed an interaction of Condition:Gender:Group at all sites. For native controls, there was an N400 effect, with Mismatch trials producing a larger N400 amplitude than Match trials, which was independent of Gender congruency across languages. For L2 learners, the N400 effect interacted with Gender congruency. The data were modeled independently thereafter for the Brazilian control group, using the same model structure as above without the fixed factor Group. For native speakers, the effect of Condition was significant at all sites due to greater mean N400 amplitude for mismatch than match trials. At frontal central sites, there was also an effect of Gender, due to a larger N400 for same gender trials. Crucially, Condition did not interact with Gender at any site (see Table 4b and Figures 4a and 4b, for same and opposite gender respectively).

Figure 4a about here

Figure 4a. *Mean voltage ERPs (and SD) for native Brazilians for 9 central electrodes as a function of Condition (Match vs. Mismatch) for nouns with same gender across languages. Permutation tests are indicated in red (p=.05) and green (p<.05) across the entire epoch.*

Figure 4a about here

Figure 4b. *Mean voltage ERPs (and SD) for native Brazilians for 9 central electrodes as a function of Condition (Match vs. Mismatch) for nouns with opposite gender across languages. Permutation tests are indicated in red (p=.05) and green (p<.05) across the entire epoch.*

Correlation between D prime score and N400 effect

Pearson's correlations were performed in order to determine whether there was any correlation between the ability to correctly identify match trials (d prime) and the magnitude of the N400 effect (cf. Tanner et al., 2013). We found no correlation between d prime scores and the magnitude of the N400 effect for either the L2 learners (same gender $r(16)=0.07$, $p = .72$, opposite gender $r(16) = 0.36$, $p = .15$) or the Brazilian control group (same gender $r(16) = 0.04$, $p = .84$, opposite gender $r(16) = 0.18$, $p = .46$). These results are depicted in Figure 5.

Figure 5 about here

Figure 5. *Regression of N400 effect as a function of D prime score for Brazilian control group and L2 learners*

Discussion

Our study revealed clear cross-linguistic gender congruency effects (GCE), from the earliest stages of acquiring a second language. This was apparent in the electrophysiological trace of lexical access, as measured by the N400. L2 learners demonstrated a clear N400 effect for mismatched visual-auditory pairs post-training, but only for newly learned nouns that shared grammatical gender across their native (French) and newly acquired language (Portuguese). No modulation of the ERP response was found as a function of the match between auditory words and visual stimuli for nouns that had opposite gender across the two languages for these learners. In addition, the effect of gender congruency was visible in the N400 modulation for match trials, for which the N400 response was increased for nouns that had opposite gender across the L2 and French compared to nouns that shared grammatical gender across languages. Hence, cross-linguistic GCE were clearly reflected in the automatic cortical response associated with lexical processing (Kutas & Federmeier, 2011). It is important to note that in the control group of native Brazilian Portuguese speakers, who on average were novice French speakers, only a robust N400 mismatch effect was found, which was independent of gender congruency. Otherwise stated, the cross-linguistic GCE revealed by the N400 was specific to the L2 learners processing nouns in the newly learned language. Importantly, this effect was found following only 4 days of training using interactive computerized games.

Previous studies on the effect of cross-linguistic gender congruency have examined this question in populations that had several years of formal learning of the second language (Bordag, 2004; Bordag & Pechmann, 2007, 2008; Costa et al., 2003; Lemhöfer et al., 2008; Morales et al., 2011; Rodriguez-Fornells & Münte, 2016; Salamoura & Williams, 2007). Our training study allowed us to examine this question from the earliest stages of acquisition. In addition, our design has the distinct advantage of presenting only the newly learned language. In several studies that have shown cross-linguistic GCE, participants had to actively process their native and second language simultaneously due to task requirements (switching between languages, translating, or ignoring embedded L1 words during L2 production; Bordag & Pechmann, 2007, 2008; Costa et al., 2003; Rodriguez-Fornells & Münte, 2016; Salamoura & Williams, 2007). The necessity to maintain both languages active may have played a role in evoking gender congruency effects in these studies. This cannot be claimed for the present results. Indeed, our study did not require L2 participants to overtly produce or consciously activate their L1. Nonetheless, that the L1 lexicon, and more specifically the grammatical features of L1 candidates, became active during L2 processing was readily apparent in the ERP data. Our results are in line with those reported by Boutonnet et al. (2012) who showed modulation of a late negative component as a function of whether triplets of words, presented in English, all shared the same lexical gender in the participants' native language, Spanish. Thus, as in the current study, even in conditions where the L1 was physically absent, it played a significant role in processing. More specifically, native speakers of "gendered" languages automatically and irrepressibly activate the L1 gender of nouns, even when processing words exclusively in the L2.

In many studies, gender congruency effects have been reported both within a given language and across languages when participants were required to produce a definite article prior to the noun (Costa et al., 2003; Salamoura & Williams, 2007);

although others have shown GCE when participants produced bare nouns (Bordag & Pechmann, 2007; Cubelli et al., 2005). In light of this, it is of interest to note that, while our participants did not produce the lexical items, all auditory nouns were preceded by the definite determiner (e.g., /oparafuzo/ "the screw" and /aluva/ "the glove"). It is an empirical question whether the gender congruency effects that we obtained would occur in the absence of the determiner. Given that our participants showed evidence that they had acquired the gender of the L2 nouns, it is possible that they may have retrieved this information during processing, either from a stored representation of the noun or from the morphology of the word form itself, which may then have been the source of interference (cf. Gollan and Frost, 2001, for a discussion of different routes to stored gender information). Further work is necessary to determine the locus of the interference we found.

Our results show rapid learning of the L2 vocabulary, as demonstrated by both the establishment of an N400 response to newly learned words and by ceiling level accuracy post-training. Concerning the cortical response, modulations of the N400 have been noted in association with L2 learning and/or artificial languages in several studies. This has been demonstrated in L2 studies that have used longitudinal designs to investigate changes in cortical activity over the course of learning (Chun, Choi, & Kim, 2012; McLaughlin, Osterhout, & Kim, 2004; Stein et al., 2006; Yum, Midgley, Holcomb & Grainger, 2014). In a seminal study, McLaughlin and colleagues (2004) found that L2 pseudowords elicited a larger N400 than learned L2 words following 14 hours of classroom instruction (McLaughlin et al., 2004). Crucially, these effects were not seen behaviorally; when making overt lexical decisions, learners performed at chance. Hence, L2 learners were sensitive to the prior exposure to word forms, as shown by the N400 effect, even when they could not consciously identify these forms.

In contrast, the semantic integration of these newly learned lexical items only occurred following 60 hours of instruction, as indexed by reduced N400 amplitude for L2 target words preceded by semantically related primes. Similarly, Soskey and colleagues (Soskey et al., 2016) reported the gradual instantiation of L2 words, as indexed by modulation of the N400, across a semester of learning. It is important to note, however, that these studies reported cortical changes due to L2 meaning integration following extended L2 training whereas we found that participants accessed the meaning of newly acquired L2 words after only 3 hours of learning.

Artificial language paradigms have been used to ascertain whether novel L2 words can be associated quickly with meaning. Breitenstein and colleagues (Breitenstein et al., 2007) used associative learning, where a spoken word was paired with the image of an object with increasing statistical probability over multiple trials and found that after 5 days of training, newly learned words facilitated (in the form of shorter response latencies) the processing of related pictures, indicating integration into existing lexical networks. A similar magnetoencephalography (MEG) study by Dobel et al. (2010) showed a reduction in the mN400 (the MEG component comparable to the ERP N400 component) to correct images preceded by trained spoken words from pre- to post-training, indicating that trained words had become associated with existing conceptual representations. Our results corroborate these findings, showing the acquisition of a small L2 vocabulary following three 25-minute training sessions and one 40-minute review session over the course of 4 days, as manifested by the establishment of an N400 response from pre- to post-training and increased accuracy, from chance to ceiling level.

Our study clearly demonstrates that learners were sensitive to the grammatical gender of the newly learned words, despite there having been no formal or explicit instruction concerning gender and even though the stimuli were only presented aurally. This implies that our participants segmented the auditory stimuli into the determiner and noun and inferred gender information from the properties of the speech signal. Segmenting the audio signal into its syntactic elements is notoriously difficult during second language acquisition (Altenberg, 2005), which is why we created a game that specifically segmented the auditory sentences and required participants to recognize the meanings of the different elements and assemble them in the correct order to recreate the auditory sentence. However, while determiner phrases and verbal phrases were explicitly segmented, the determiner phrases were presented as a single unit (e.g., /okaSimbo/ "the pipe" and /asaia/ "the skirt"). Learners could, in theory, have interpreted the determiner phrase as a whole rather than segment it into the determiner and noun, as indeed there is evidence for in young children. For example, in early stages of acquisition French children may produce forms that reveal segmentation errors (e.g., "*le loiseau*" and "*le noiseau*" stemming from the speech signal "*l'oiseau*" and "*un oiseau*" «the bird / a bird» Clark, 2009). Even so, the regularity of the morphological form for the determiner (/a/ or /o/) preceding the noun and the concurring final phoneme of nouns reduces this possibility, as discussed below. Moreover, the majority of participants had formally learned Spanish as a third language throughout secondary school, which may well have prompted them to capitalize on the partial overlap of gender concord rules in Spanish and Portuguese (cf. Brooks & Kempe, 2013).

Various studies with either natural or artificial languages have shown that following implicit training, adult learners rapidly deduce the rules that govern grammatical gender assignment (Öttl & Behne, 2017) and gender concord (Denhovska & Serratrice, 2017; Morgan-Short et al., 2010). Concerning gender assignment, our materials provided a clear phonetic cue, as outlined above, but no semantic information was associated with the gender of nouns. This differs from the materials learned in an artificial language (Öttl & Behne, 2017) in which gender suffixes on the noun were determined by the biological gender of stimuli. Concerning gender concord, Morgan-Short et al. (2010) found no difference in learning as measured by behavioral (d prime) or cortical sensitivity (P600 response) to determiner-noun gender concord violations as a function of the type of training (implicit or explicit) at the end of training. Using a miniature set of Russian nouns and adjectives, in which adult learners were exposed to noun-adjective gender concord in short written sentences, Denhovska and Seratrice (2017) showed that even under implicit learning, where emphasis was placed on learning the meanings of the sentences and no mention was made of the underlying grammatical rules, participants readily acquired these rules. Moreover, no difference in behavioral accuracy was found for grammatical judgments as a function of the type of instruction (implicit or explicit), although only those who received explicit instruction were able to produce the grammatical rules governing gender concord above chance level. The present results are in line with those found in the above studies, showing that participants rapidly acquire grammatical gender concord rules in a newly learned language following short training sessions, even in the absence of formal instruction.

The pattern of results we obtained suggests that gender congruency effects play an early role during lexical access. This question, i.e. whether grammatical gender exerts an early influence on lexical access or only later, during lexical selection, has been examined in various monolingual studies. Eye-tracking experiments have demonstrated that both children and adults use gender agreement to predict nouns when they are preceded by a gender-marked determiner (Brouwer et al., 2017; Cholewa et al., 2019; Lew-Williams & Fernald, 2007). Far less evidence of this has been found in the second language (Hopp & Lemmerth, 2016; Lemmerth & Hopp, 2019). In a primed lexical decision task using auditory homophone primes and orthographic targets, Spinelli and Alario (2002) found that gender-marked determiners constrained lexical access to the gender compatible candidate for French homophones. However, subsequent work provided evidence that grammatical gender does not in fact constrain lexical access but acts at a later stage during the selection of the appropriate candidate (Spinelli et al., 2006). Our results clearly demonstrate that the L1 gender of stored words was activated and interacted with the L2 gender of actually presented words. It remains to be determined whether such was due to the presence of the salient and reliable morphological marking carried in the determiner (i.e. whether participants computed gender based on morpho-phonological cues) or due to the activation of the L2 gender from the auditory noun itself (i.e. retrieval of gender from a newly stored representation). Nevertheless, our results suggest that gender congruency across languages affected lexical access for newly learned L2 nouns, which was hindered when competing gender features from the L1 were activated.

Our results, showing a clear interaction between the established L1 and a newly acquired L2 vocabulary, add to the ample body of experimental evidence showing that the interaction between languages is ubiquitous and found at all levels of processing (see Kroll & Bialystok, 2013, for a review of experimental studies). Indeed, the automatic activation across languages has been demonstrated during both auditory and written processing at the phonological level (Carrasco et al., 2012; Friesenet al., 2020), the lexical level (Dijkstra et al., 2000; Lagrou et al. , 2011; Sunderman & Kroll, 2006) and the syntactic level (Dussias & Sagarra, 2007; Foucart & Frenck-Mestre, 2011). Moreover, parallel activation across languages is found despite distinct writing systems (Thierry & Wu, 2012; Wu & Thierry, 2010). Our results present further evidence that even in monolingual contexts and even for novice learners, the two languages are activated in parallel.

Differences between ERP and behavioral results are likely due to discrepancies in timing and granularity. Whereas the N400 is time-locked to stimuli in such a way as to provide information regarding early aspects of lexical-semantic processing, behavioral measures can encompass both early and late effects during processing. Otherwise stated, by the time low-temporal resolution behavioral data, such as accuracy or reaction time, are measured, other, later, cognitive processes have plausibly occurred (ex. working and episodic memory processes, response selection, mental imagery, c.f. Hauk, 2016). Specifically pertaining to gender congruency effects during a semantic categorization task in the L2, previous ERP results showed a negative (LAN) modulation for cross-language gender incongruent words but no differences in reaction time between conditions (Boutonnet et al., 2012). Similarly, in our study, an N400 effect was seen for Match vs. Mismatch trials but only for cross-language gender congruent words, suggesting that cross-language gender congruency had a direct effect on semantic processing. Once again, this effect was not seen in behavioral results (d prime or accuracy). We posit that in both our and Boutonnet et al.'s (2012) studies, cross-language gender interference effects happen too early in lexical processing to be evidenced behaviorally, which explains why the GCE as measured by EEG (N400) was not echoed by a behavioral effect. Our results are also in line with those reported by McLaughlin and colleagues (2004) who found cortical sensitivity to the lexical status of newly learned L2 items in the absence of behavioral capacity to distinguish between words and non-words. In similar fashion our results show cortical sensitivity to gender congruency across languages that was not reflected in the behavioral response.

Last, we can note that our design did not allow us to examine the effect of gender concord or gender congruency within the L2. Several studies are warranted to delve further into this question. It would be of interest, for example, to test whether we could establish gender congruency effects within the newly learned L2 by adopting a paradigm similar to that used by Boutonnet et al. (2012), in which successive trials carried the same L2 gender, or a visual world paradigm (Hopp & Lemmerth, 2016) in which the gender of the items is varied both across languages and within the L2. Concerning gender concord, a typical violation paradigm could be added to the current design, whereby the learned L2 nouns are preceded by either the correct determiner or the incorrect determiner (Foucart & Frenck-Mestre, 2011, 2012; Morgan-Short et al., 2010). We demonstrated previously that the overlap of grammatical features between the learners' L1 and L2, the specific syntactic structure and the level of proficiency in the L2 all play a role in the pattern of ERP components elicited by gender concord violations in sentential contexts for L2 speakers (Foucart & Frenck-Mestre, 2011, 2012). In similar fashion, recent work that manipulated gender concord in sentential contexts found that the ERP signature for these violations – generally the LAN/P600 – can vary even in native speakers as a function of stimulus characteristics (Beatty- Martínez, et al., 2021). This challenges the notion that ERP signatures associated with syntactic processing (error detection) are tied to a specific level of processing and/or are more systematic in native speakers than in L2 learners. Hence, in novice learners of a gendered language, whether we would find an ERP response to gender concord errors within the newly learned language and, if so, what component would be elicited is an open question.

To conclude, we have provided clear electrophysiological evidence of gender congruency effects in an L2, from the very beginning of acquisition. The clear N400 effect elicited by mismatched compared to matched audio-visual pairs for gender congruent trials was basically annulled for gender incongruent trials. To our knowledge, no prior work has provided evidence of GCE, either behaviorally or via the cortical response to newly learned L2 words. It is important to note that our behavioral results clearly demonstrate that participants learned the correct association between auditory words and images and that this was independent of both the gender in the second language vocabulary and, importantly, of the congruency of gender across the L2 and the participants' L1. Hence, we have also provided evidence of the importance of a multi-disciplinary approach to bring the effects of cross-linguistic gender congruency to light, at least in the early stages of acquisition.

Tables.

Fixed effects: Midline Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.24420	0.53291	2.335	0.0328*
COND.sum1	0.43901	0.20430	2.149	0.0442*
GEND.sum1	-0.19069	0.10448	-1.825	0.0680
EXPE.sum1	0.04772	0.10497	0.455	0.6494
COND.sum1:GEND.sum1	-0.30688	0.13645	-2.249	0.0338*
GEND.sum1:EXPE.sum1	-0.10439	0.10450	-0.999	0.3179
COND.sum1:EXPE.sum1	0.47717	0.10474	4.556	5.31e-06 ***
COND.sum1:GEND.sum1:EXPE.sum1	-0.42132	0.10454	-4.030	5.63e-05 ***
Fixed effects: Frontal Central Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.191e-01	4.801e-01	1.915	0.07348
COND.sum1	3.745e-01	1.859e-01	2.015	0.06091
GEND.sum1	-2.213e-01	6.605e-02	-3.350	0.00255 **
EXPE.sum1	5.209e-02	6.349e-02	0.820	0.41200
COND.sum1:GEND.sum1	-2.581e-01	6.322e-02	-4.082	4.48e-05
GEND.sum1:EXPE.sum1	-1.010e-01	6.325e-02	-1.597	0.11020
COND.sum1:EXPE.sum1	4.016e-01	6.344e-02	6.331	2.50e-10**
COND.sum1:GEND.sum1:EXPE.sum1	-2.057e-01	6.326e-02	-3.251	0.00115 **
Fixed effects: Central Parietal Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.525e-01	4.365e-01	1.266	0.224
COND.sum1	4.020e-01	1.799e-01	2.234	0.040 *
GEND.sum1	-7.605e-02	6.285e-02	-1.210	0.226
EXPE.sum1	-8.894e-03	6.317e-02	-0.141	0.888
COND.sum1:GEND.sum1	-2.463e-01	6.285e-02	-3.919	8.92e-05***
GEND.sum1:EXPE.sum1	-3.943e-02	6.286e-02	-0.627	0.530
COND.sum1:EXPE.sum1	4.686e-01	6.312e-02	7.424	1.19e-13**
COND.sum1:GEND.sum1:EXPE.sum1	-2.634e-01	6.287e-02	-4.189	2.81e-05**

Model: lmer(MVC ~(1+COND.sum|SUBJECTS) + (1+COND.sum|ITEM) + COND.sum + GEND.sum + EXPE.sum + COND.sum:GEND.sum + GEND.sum:EXPE.sum + COND.sum:EXPE.sum + GEND.sum:EXPE.sum:COND.sum)

Table 3a. *Model output for mean voltage ERPs for L2 learners as a function of Session (Pre vs. Post), Gender (Same vs. Opposite) and Condition (Match vs. Mismatch).*

Fixed effects Midline Sites	Estimate	Std. Error	t value	Pr(>|t|
(Intercept)	1.20538	0.58093	2.075	0.0533
COND.sum1	-0.02400	0.22892	0.105	0.9177
GEND.sum1	-0.09297	0.17807	0.522	0.6061
COND.sum1:GEND.sum1	0.06651	0.14335	0.464	0.6427

Fixed effects Frontal Central Sites	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.775e-01	5.270e-01	1.665	0.115
COND.sum1	7.535e-03	1.714e-01	0.044	0.965
GEND.sum1	-1.183e-01	8.777e-02	-1.347	0.178
COND.sum1:GEND.sum1	-8.305e-02	8.783e-02	-0.946	0.344

Fixed effects Central Parietal Sites	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.58026	0.44970	1.290	0.215
COND.sum1	-0.05230	0.20715	-0.252	0.804
GEND.sum1	-0.03018	0.08599	-0.351	0.726
COND.sum1:GEND.sum1	-0.01232	0.08601	-0.143	0.886

Model: lmer(MVC ~(1+COND.sum|SUBJECTS) + (1+COND|ITEM) + COND.sum + GEND.sum + COND.sum:GEND.sum, data = d)

Table 3b. *Model output for mean voltage ERPs for L2 learners pre-training as a function of Gender (Same vs. Opposite) and Condition (Match vs. Mismatch).*

Fixed effects: Midline				
	Estimate	Std. Er	t value	Pr(>|t|)
(Intercept)	0.4832	0.5397	1.117	0.28032
COND.sum1	0.8596	0.2227	3.859	0.00133**
GEND.sum1	-0.1175	0.0911	1.289	0.19730
COND.sum1:GEND.sum1	-0.5198	0.0911	-5.705	0.00000001
Fixed effects:Frontal Central				
	Estimate	Std. Er	t value	Pr(>|t|)
(Intercept)	0.95444	0.53177	1.795	0.09117
COND.sum1	0.74904	0.24744	3.027	0.00779**
GEND.sum1	-0.3247	0.10393	-3.125	0.00433**
COND.sum1:GEND.sum1	-0.48185	0.09068	-5.314	0.00000013
Fixed effects: Central Parietal				
	Estimate	Std. Er	t value	Pr(>|t|)
(Intercept)	0.5397	0.4832	1.117	0.28032
COND.sum1	0.8596	0.2227	3.859	0.00133 **
GEND.sum1	-0.1175	0.0911	-1.289	0.19730
COND.sum1:GEND.sum1	-0.5198	0.0911	-5.705	0.00000001

Model: lmer(MVC ~(1+COND.sum|SUBJECTS) + (1+ COND.sum |ITEM) + COND.sum + GEND.sum + COND.sum:GEND.sum)

Table 3c. *Model output for mean voltage ERPs for L2 learners post-training as a function of Gender (Same vs. Opposite) and Condition (Match vs. Mismatch)*

Simple effect of CONDITION for Same Gender trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.2598	0.6633	4.915	0.000146***
CONDmis	-3.4028	0.7703	-4.417	0.000525***
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.559	0.7118	3.596	0.00225
CONDmis	-2.6034	0.6816	-3.820	0.00151**
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.0111	0.5935	3.388	0.003669 **
CONDmis	-2.9107	0.5770	-5.045	0.000127***
Simple effect of CONDITION for Opposite Gender trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.0730	0.8901	1.205	0.244
CONDmis	-0.2687	0.7569	-0.355	0.727
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8607	0.8055	1.069	0.301
CONDmis	-0.5241	0.6441	-0.814	0.427
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.7543	0.7770	0.971	0.346
CONDmis	-0.5932	0.6129	-0.968	0.347

Model: lmer(MVC ~(1+COND|SUBJECTS) + (1+ COND|ITEM) + COND)

Table 3d. *Model output for mean voltage ERPs for L2 learners post-training as a function of Condition (Match vs. Mismatch) for nouns of the opposite (above) and same gender (below) across languages.*

Simple effect of Gender for Match trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.1568	0.7233	1.599	0.126
GENDsame	2.0753	0.4222	4.916	9.62e-07
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8604	0.8033	1.071	0.3000
GENDsame	1.6908	0.5854	2.888	0.0107
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.7621	0.6029	1.264	0.223
GENDsame	1.2745	0.2570	4.960	7.19e-07***
Simple effect of Gender for Mismatch trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8051	0.5736	1.404	0.1754
GENDsame	-0.9082	0.4340	-2.093	0.0365*
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.3356	0.5399	0.621	0.543
GENDsame	-0.3838	0.8047	-0.477	0.640
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1590	0.7017	0.227	0.824
GENDsame	1.0637	0.8916	-1.193	0.250

Model: lmer(MVC ~(1+GEND|SUBJECTS) + (1+ GEND|ITEM) + GEND)

Table 3e. *Model output for mean voltage ERPs as a function of Gender (Same vs. Opposite) in the L2 learner group*

Model summary at midline sites				
Fixed effects:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.0468	0.4197	2.494	0.01758
COND.sum1	1.1548	0.2038	5.665	2.59e-06
GEND.sum1	0.1021	0.1255	0.814	0.42275
GROUP.sum1	0.2257	0.4148	0.544	0.59009
COND.sum1:GEND.sum1	0.3167	0.1127	-2.811	0.00626
GEND.sum1:GROUP.sum1	0.1904	0.1081	1.760	0.07838
COND.sum1:GROUP.sum1	0.2434	0.2014	1.209	0.23548
COND.sum1:GEND.sum1:GROUP.sum1	0.4298	0.1081	3.975	7.11e-05
Model summary at frontal central sites				
Fixed effects:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.075e-01	4.017e-01	2.010	0.052510
COND.sum1	9.765e-01	1.872e-01	5.218	9.42e-06***
GEND.sum1	3.382e-01	7.398e-02	4.572	0.000102***
GROUP.sum1	1.450e-01	4.003e-01	0.362	0.719407
COND.sum1:GEND.sum1	2.782e-01	6.611e-02	-4.208	2.80e-05***
GEND.sum1:GROUP.sum1	-1.400e-02	6.585e-02	-0.213	0.831647
COND.sum1:GROUP.sum1	2.248e-01	1.871e-01	1.202	0.237987
COND.sum1:GEND.sumsame:GROUP.sum1	2.039e-01	6.585e-02	3.096	0.001966**
Model summary at central parietal sites				
Fixed effects:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.025e-01	3.656e-01	0.554	0.58341
COND.sum1	1.237e+00	2.034e-01	6.083	7.26e-07***
GEND.sum1	-1.375e-01	6.475e-02	-2.124	0.03368 *
GROUP.sum1	-3.365e-01	3.656e-01	-0.920	0.36398
COND.sum1:GEND.sum1	-2.873e-01	6.476e-02	-4.436	9.21e-06***
GEND.sum1:GROUP.sum1	-1.978e-02	6.475e-02	0.305	0.76001
COND.sum1:GROUP.sum1	3.761e-01	2.034e-01	1.849	0.07329
COND.sum1:GEND.sum1:GROUP.sum1	2.331e-01	6.476e-02	3.599	0.00032***

Model: lmer(MVC ~(1+COND.sum|PARTICIPANTS) + (1+COND.sum|ITEM) + COND.sum:GEND.sum + GEND.sum:GROUP.sum + COND.sum:GROUP.sum+ GEND.sum:GROUP.sum:COND.sum, data = d)

Table 4a. *Model output for mean voltage ERPs (post-training for L2 learners) as a function of Group (Native vs. L2), Gender (Same vs. Opposite) and Condition (Match vs. Mismatch)*

Model summary: midline sites				
Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.82079	0.60273	1.362	0.190656
COND.sum1	1.39788	0.31593	4.425	0.000387 ***
GEND.sum1	0.08912	0.16977	0.525	0.603694
COND.sum1:GEND.sum1	0.11428	0.15576	0.734	0.463721
Model summary: frontal central sites				
Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.66200	0.59535	1.112	0.281627
COND.sum1	1.20152	0.27855	4.313	0.000475 ***
GEND.sum1	-0.35212	0.09536	3.693	0.000223 ***
COND.sum1:GEND.sum1	-0.07351	0.09536	0.771	0.440817
Model summary: central parietal sites				
Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.13409	0.54457	-0.246	0.808454
COND.sum1	1.61456	0.33473	4.823	0.000157***
GEND.sum1	-0.15744	0.09181	-1.715	0.086405
COND.sum1:GEND.sum1	0.05374	0.09181	0.585	0.558323

Model: lmer(MVC ~(1+COND.sum|PARTICIPANTS) + (1+COND.sum|ITEM) + COND.sum:GEND.sum + GEND.sum:COND.sum, data = d)

Table 4b. . *Model output for mean voltage ERPS for the native control group as a function of Gender (Same vs. Opposite) and Condition (Match vs. Mismatch)*

References

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531–552.

Alencar de Resende N.C., Mota M.B., Seuren P. (2019) The Processing of Grammatical Gender Agreement in Brazilian Portuguese: ERP Evidence in Favor of a Single Route. *J Psycholinguist Res. 48*(1),181-198.

Altenberg, E.P. (2005). The perception of word boundaries in a second language. *Second Language Research*, *21*, 325-358.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305.

Barber, H., & Carreiras, M. (2003). Integrating Gender and Number Information in Spanish Word Pairs: An Erp Study. *Cortex*, *39*(3), 465–482.

Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. Journal of Cognitive Neuroscience, *17*, 137-155.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Beatty-Martínez, A. L., Bruni, M. R., Bajo, M. T., & Dussias, P. E. (2021). Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers. *Psychophysiology*, *58*(30).

Bigi, B. (2015). SPPAS - Multi-lingual approaches to the automatic annotation of speech. *The Phonetician*, *111– 112*(ISSN:0741-6164), 54–69.

Boersma, P. & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/

Bordag, D. (2004). Interaction of L1 and L2 systems at the level of grammatical encoding: Evidence from picture naming. *EUROSLA Yearbook*, *4*(1), 203–230.

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition, 10*, 299–314.

Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research, 24*, 139–166.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a Neurobiologically Plausible Model of Language-Related, Negative Event-Related Potentials. *Frontiers in Psychology*, *10*, 298.

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79

Breitenstein, C., Zwitserlood, P., de Vries, M., Feldhues, C., Knecht, S., & Dobel, C. (2007). Five days versus a lifetime: Intense associative vocabulary training generates lexically integrated words. *Restorative Neurology and Neuroscience*, *25*(5,6), 493–500.

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition*, *41*(2), 281–296.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science, 41*(Suppl 6), 1318–1352.

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: evidence from eye movements *Journal of Experimental Child Psychology*, *159*, 50–65.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177-208.

Carrasco, H., Midgley, K. & Frenck-Mestre, C. (2012). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones. *Psychophysiology*, *49*, 531-543.

Cholewa, J., Neitzle, I., Bürsgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: An eye-tracking study with monolingual German 3rd and 4th graders. *Frontiers in Psychology*, *10*, Article 2586.

Chun, E., Choi, S., & Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study. *Neuroscience Letters*, *521*(2), 125–129.

Clark, E.V. (2009). *First Language Acquisition*. 2nd Edition. Cambridge, UK: Cambridge University Press.

Corrêa L, Augusto M. & Castro A., (2011). Agreement and markedness in the ascription of gender to novel animate nouns by children acquiring Portuguese. *Journal of Portuguese Linguistics*, *10*(1), 121-142.

Corréa, L.M.S., & Name, M.C. (2003). The processing of determiner-noun agreement and the identification of the gender of nouns in the early acquisition of Portuguese. *Journal of Portuguese Linguistics*, *2*, 19–43.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296.

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181–200.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M. & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, *53*, 42–59.

Dahan, D., Swingley, D., Tanenhaus, M., & Magnuson, J. (2000). Linguistic Gender and Spoken-Word Recognition in French. *Journal of Memory and Language*, *42*, 465–480.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition, 135,* Article 103569. doi.org/10.1016/j.bandc.2019.05.007

Denhovska, N., & Serratrice, L. (2017). Incidental Learning of Gender Agreement in L2. *Journal of Psycholinguistic Research*, *46*(5), 1187–1211.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On Being Blinded by Your Other Language: Effects of Task Demands on Interlingual Homograph Recognition. *Journal of Memory and Language, 42*(4), 445-464.

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: On the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*(6), 1251–1261.

Dowens, M. G., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish—Evidence from Event Related Potentials. *Neuropsychologia*, *49*(7), 1651–1659.

Dussias, P.E, & Sagarra N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101–116.

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in romance languages: Evidence from French. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1414–1421.

Foucart, A., & Frenck−Mestre, C. (2012). Can late learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*, 226-248.

Foucart, A., & Frenck−Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1 -L2 syntactic similarity. *Bilingualism, Language and Cognition*, *14*, 379-399.

Fraga, I., Padrón & Hinojosa, J. A. (2021) Negative valence effects on the processing of agreement dependencies are mediated by ERP individual differences in morphosyntactic processing, *Language, Cognition and Neuroscience*

Enriquez-Guppert, S., Komrad, C., Pantev, C., & Huster, R.J. (2010) Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *NeuroImage*, *51*, 877-887.

Frenck−Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research*, *21*, 175-198.

Friesen, D. C., Ward, O., Bohnet, J., Cormier, P., & Jared, D. (2020). Early activation of cross-language meaning from phonology during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1754–1767.

Gollan, T., & Frost, R. (2001). Two Routes to Grammatical Gender: Evidence from Hebrew. *Journal of Psycholinguistic Research*, *30*, 627-651

Gunter, T., Friederici, A., & Schriefers, H. (2000). Syntactic Gender and Semantic Expectancy: ERPs Reveal Early Autonomy and Late Interaction. *Journal of Cognitive Neuroscience*, *12*, 556-568.

Hagoort, P. (2003). Interplay between Syntax and Semantics during Sentence Comprehension: ERP Effects of Combining Syntactic and Semantic Violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61.

Hopp, H., & Lemmerth, N. (2016). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*, 1-29.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge, UK: Cambridge University Press.

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*, 9–28.

Kroll, J.F. & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. Journal of Cognitive Psychology (Hove), 25(5):

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205–225.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences:Brain potentials reflect semantic incongruity. *Science, 207,* 203-205.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26.

Lagrou E., Hartsuiker R.J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 952–965.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*(3), 312–330.

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross linguistic lexical and syntactic influences. *Language Acquisition*, *26*(1), 21–45.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38.

Lew-Williams, C., & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition. *Psychological Science*, *18*(3), 193–198.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704.

Meltzer, J., & Braun, A. (2013). P600-like positivity and Left Anterior Negativity responses are elicited by semantic reversibility in nonanomalous sentences. *Journal of Neurolinguistics*, *26*, 129–148.

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement: a reply to Tanner et al. *Cortex*, *66*, 156–159.

Morales, L., Paolieri, D., & Bajo, T. (2011). Grammatical Gender Inhibition in Bilinguals. *Frontiers in Psychology*, *2*. doi.org/10.3389/fpsyg.2011.00284

Morales, L., Paolieri, D., Dussias, P. E., Valdés Kroff, J. R., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism*, *19*(2), 294–310.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, *60*(1), 154–193.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524.

Nichols, E. S., & Joanisse, M. F. (2016). Functional activity and white matter microstructure reveal the independent effects of age of acquisition and proficiency on second-language learning. *NeuroImage*, *143*, 15–25.

Öttl, A., & Behne, D. M. (2017). Assessing the Formation of Experience-Based Gender Expectations in an Implicit Learning Scenario. *Frontiers in Psychology*, *8*. doi.org/10.3389/fpsyg.2017.01485

Paolieri, D., Cubelli R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *Quarterly Journal of Experimental Psychology, 63*, 1631-1645.

Paolieri, D., Demestre, J., Guasch, M., Bajo, T. & Ferré, P. (2020) The gender congruency effect in Catalan-Spanish bilinguals: Behavioral and electrophysiological evidence. *Biingualism, Language and Cognition*, 23, 1045-1055.

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language, 18*, 571–590.

Popov, S. & Bastiaanse, R. (2018). Processes underpinning gender and number disagreement in Dutch: An ERP study. *Journal of Neurolinguistics*, *46*, 109-121.

Popov, S., Miceli, G., Ćurčić-Blake, B., & Bastiaanse, R. (2020). The role of semantics and repair processes in article noun gender disagreement in Italian: An ERP study. *Brain and Language*, *206*, 104787.

Rodriguez-Fornells, A., & Münte, T.F. (2016). Syntactic interference in bilingual naming during language switching: An electrophysiological study. In J.W. Schweiter (Ed.) *Cognitive Control and Consequences of Multilingualism*, John Benjamins, Amsterdam/Philidelphia. pp. 239-270.

Sá-Leite, A. R., Fraga, I., & Comesaña, M. (2019). Grammatical gender processing in bilinguals: An analytic review. *Psychonomic Bulletin & Review, 26*(4), 1148–1173.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are L1 and L2 processed similarly. *Second Language Research*, *24*, 397–430.

Sabourin, L., Stowe, L.A., & de Haan, G.J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*, 1–29

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*(3), 257–275.

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., & Martín-Loeches, M. (2014). Differential Task Effects on N400 and P600 Elicited by Semantic and Syntactic Violations. *PLoS ONE*, *9*(3), e91226.

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*(1), 169–194.

Seigneuric, A., Zagar, D., Meunier, F. & Spinelli, S. (2007). The relation between language and cognition in children aged 3 to 9 : The acquisition of grammatical gender in French. *Journal of Experimental Child Psychology*, 96, 229-246.

Soskey, L., Holcomb, P. J., & Midgley, K. J. (2016). Language effects in second-language learners: A longitudinal electrophysiological study of Spanish classroom learning. *Brain Research*, *1646*, 44–52.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Spinelli, E., & Alario, F.-X. (2002). Gender Context Effects on Homophone Words. *Language and Cognitive Processes*, 457–469.

Spinelli, E., Meunier, F., & Seigneuric, A. (2006). Spoken word recognition with gender-marked context. *The Mental Lexicon*, 277–297.

Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *NeuroImage*, *33*(2), 774–783.

Sunderman G., & Kroll, J.F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, 28, 387–422.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex*, *66*, 149–155.

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367–382.

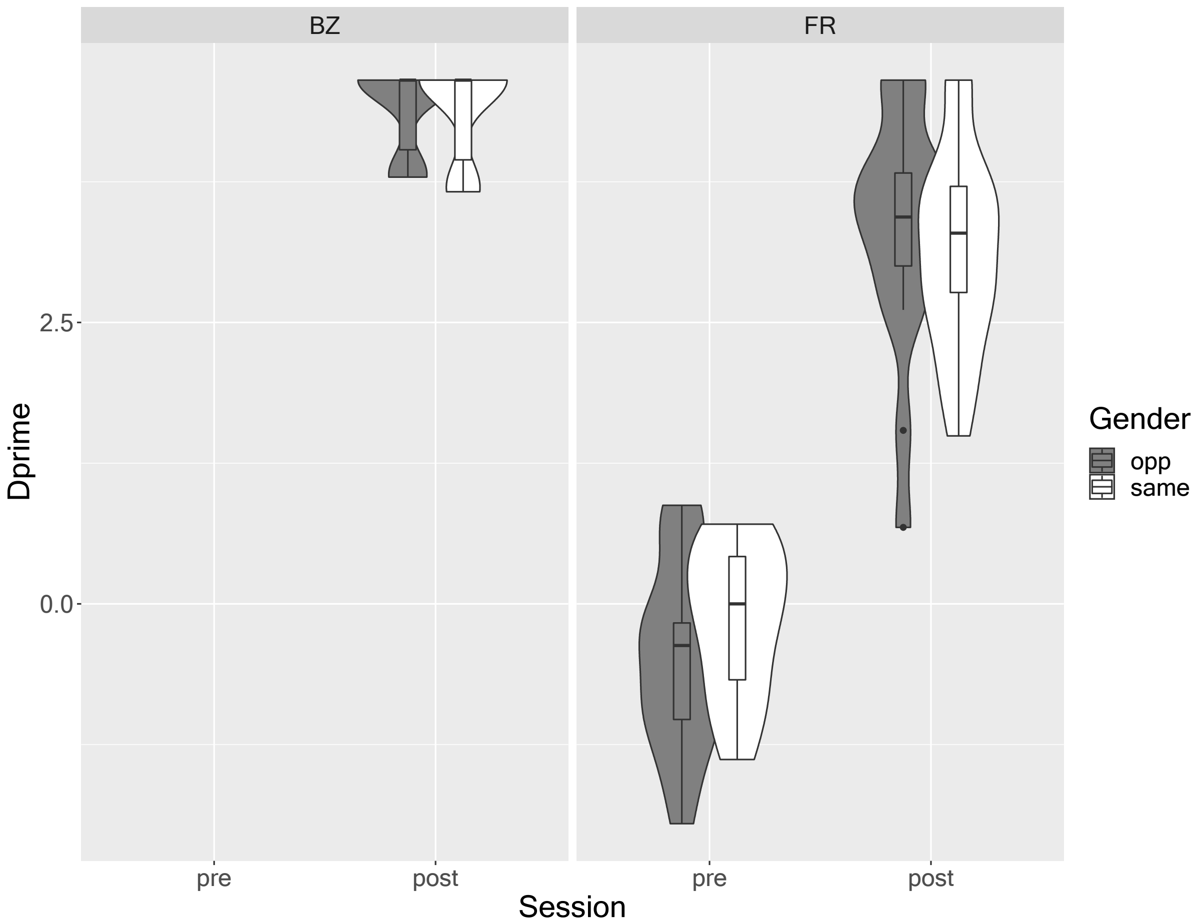
Thierry, G., & Wu, Y. (2017). Brain potentials reveal unconscious translation during foreign-language comprehension. Proceedings of the National Academy of Science, *104*: 12530±12535

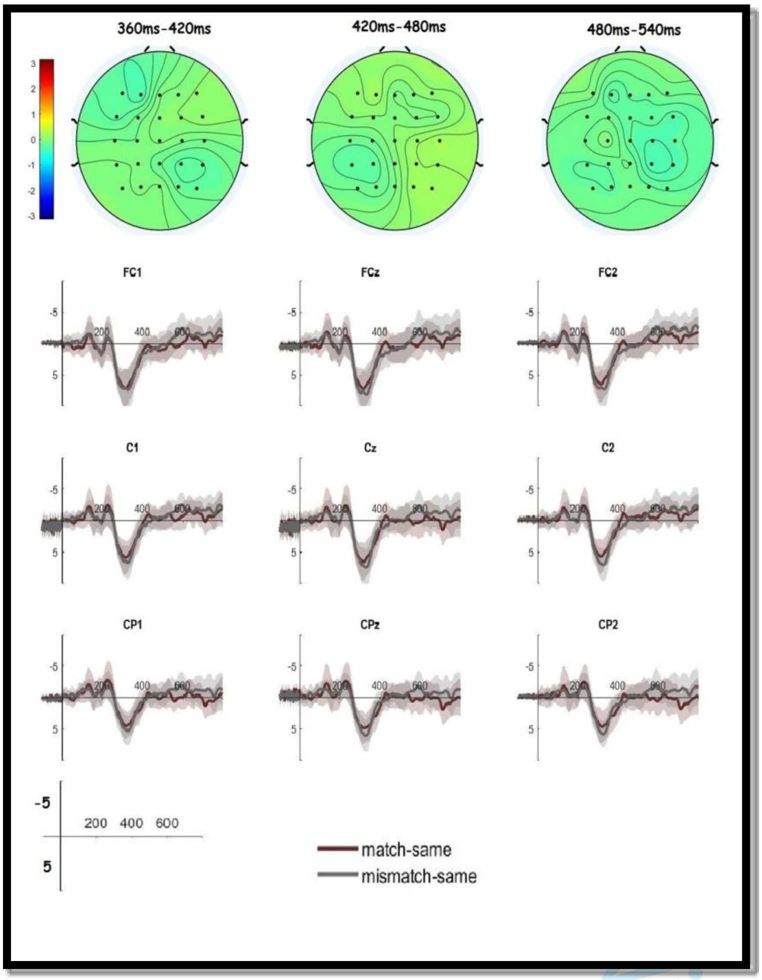
Tokowicz, N., & MacWhinney, B. (2005). Implicit and Explicit Measures of Sensitivity to Violations in Second Language Grammar: An Event-Related Potential Investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204.

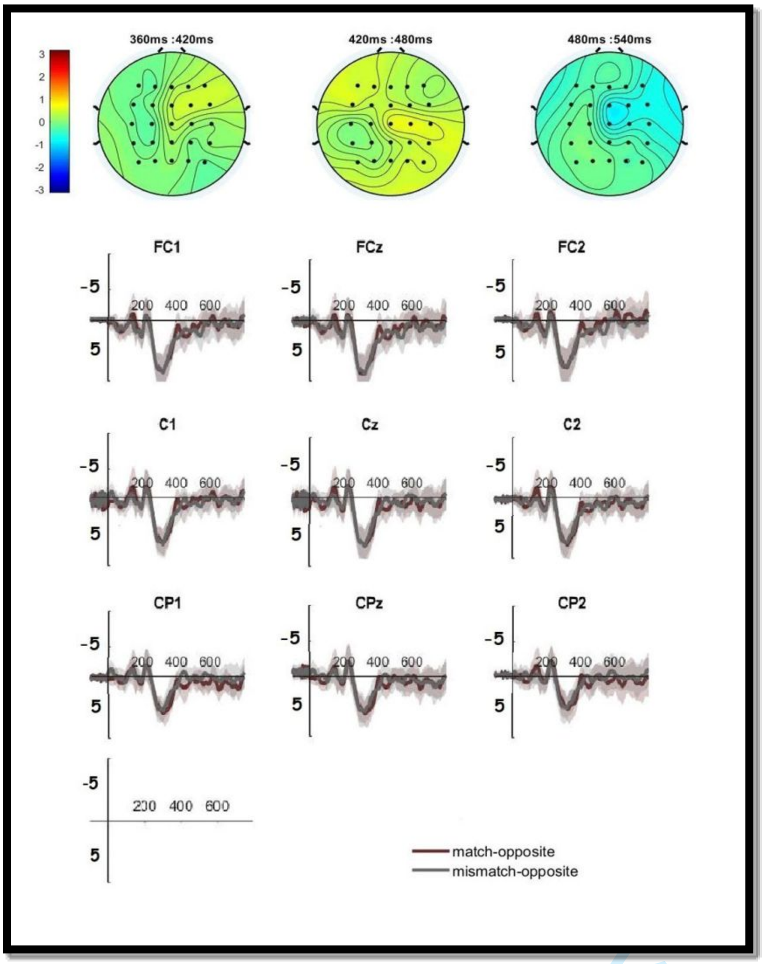
Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288.

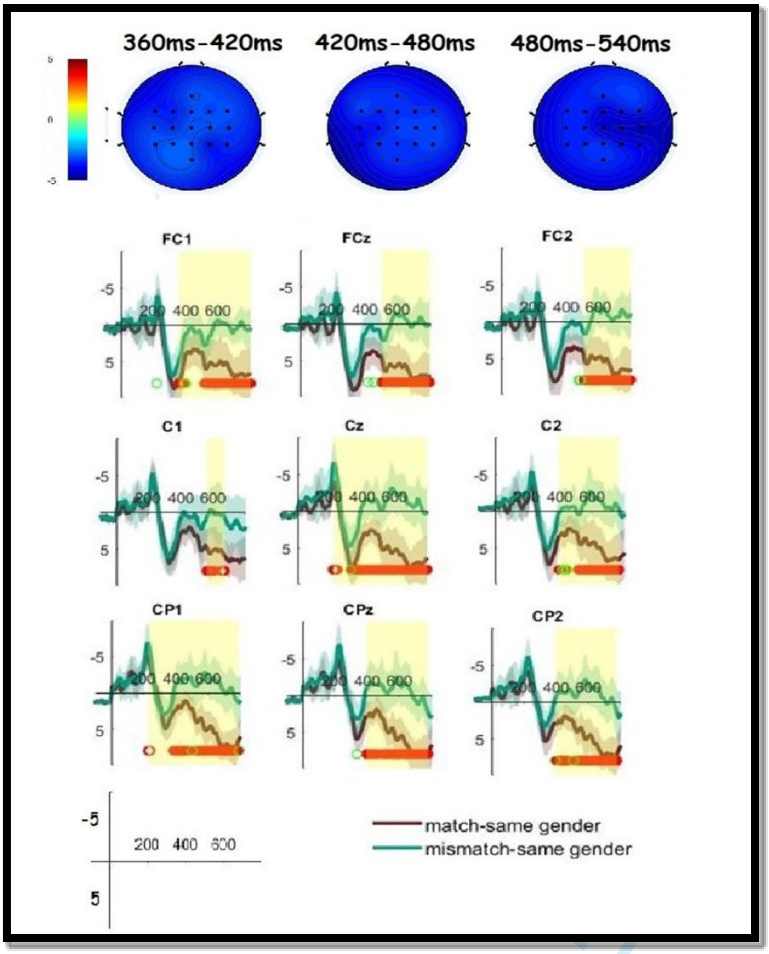
Wu., X & Thierry, G. (2010). Chinese–English Bilinguals Reading English Hear Chinese (2010). *The Journal of Neuroscience*, 30, 7646 –7651

Yum, Y., Midgley, K., Holcomb, P., & Grainger, J. (2014). An ERP study on initial second language vocabulary learning. *Psychophysiology*, 51(4), 364–373.

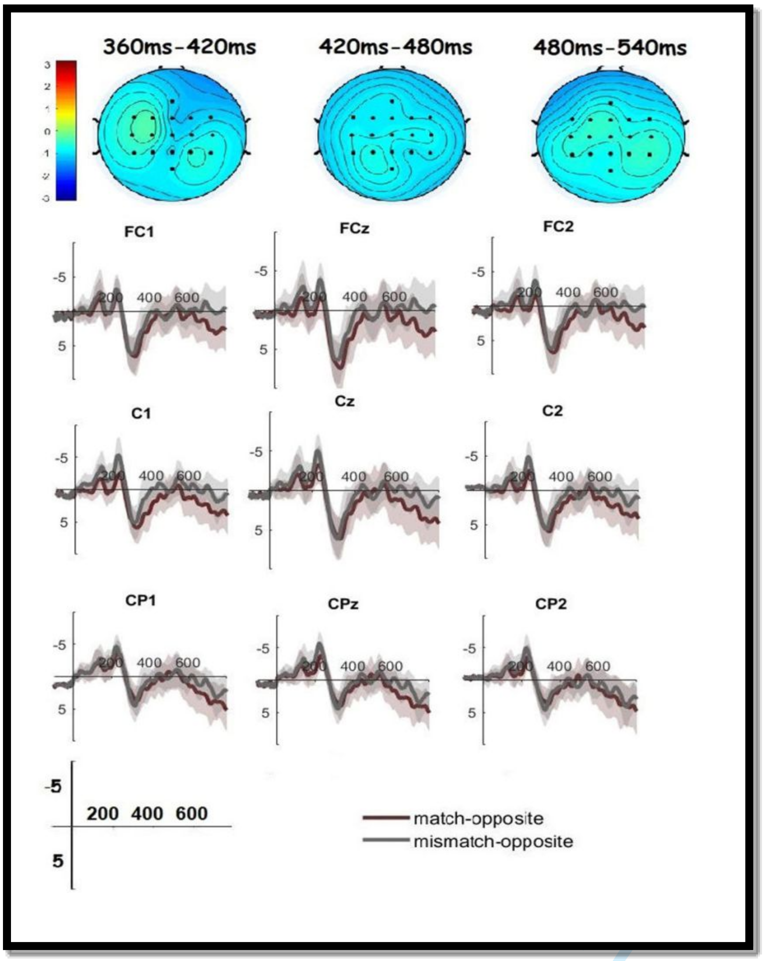

BZ
FR
2.5
0.0
Dprime
pre
post
pre
post
Session
Gender
opp
same

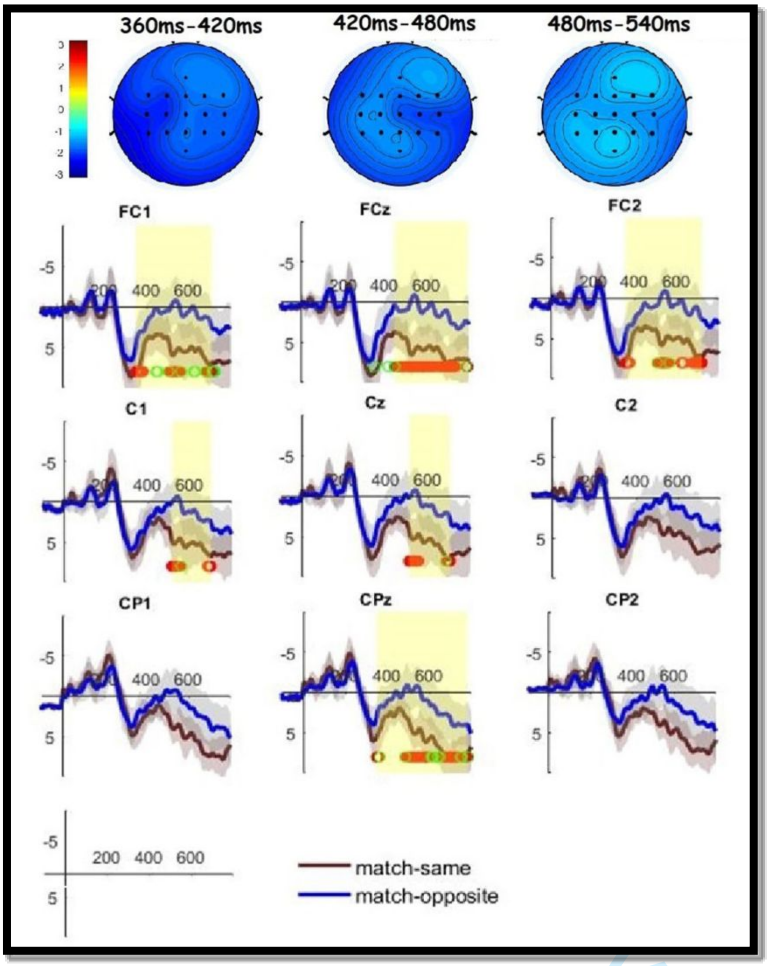

360ms-420ms
420ms-480ms
480ms-540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
-5
200 400 600
5
match-same
mismatch-same


360ms :420ms
420ms :480ms
480ms :540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
match-opposite
mismatch-opposite

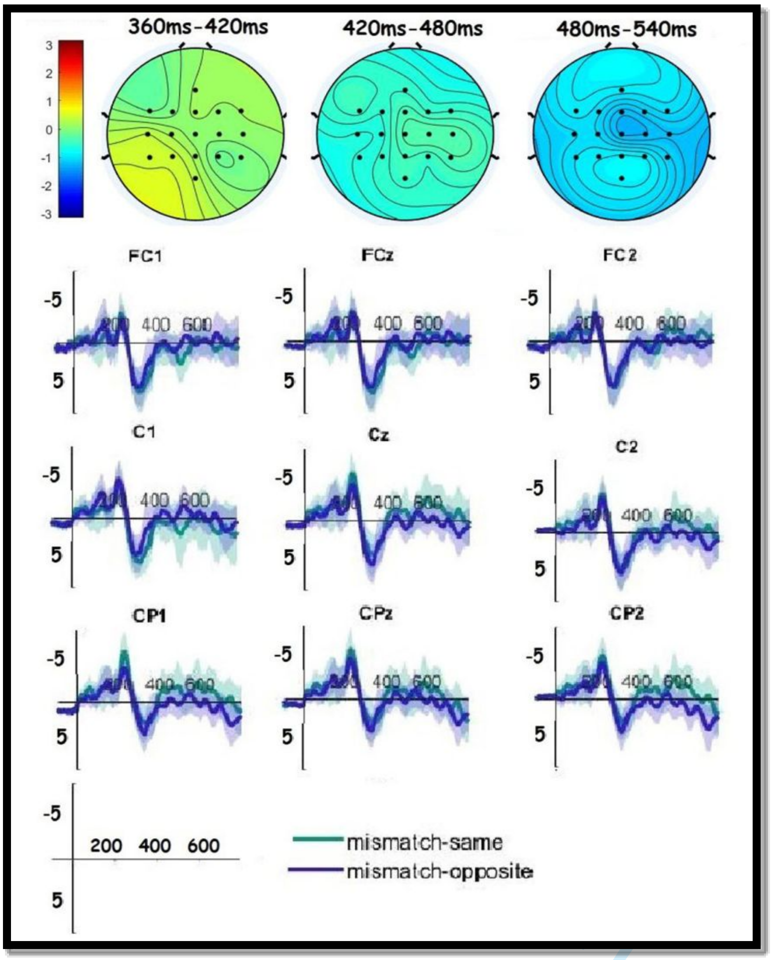


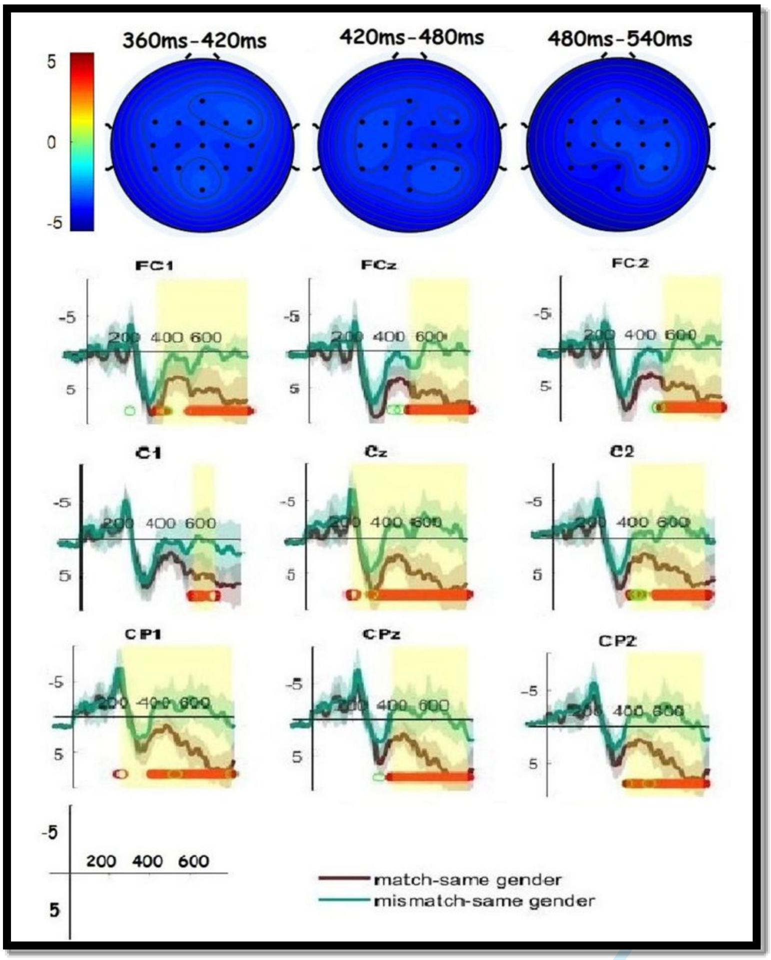
360ms-420ms
420ms-480ms
480ms-540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
-5
200 400 600
5
match-same gender
mismatch-same gender

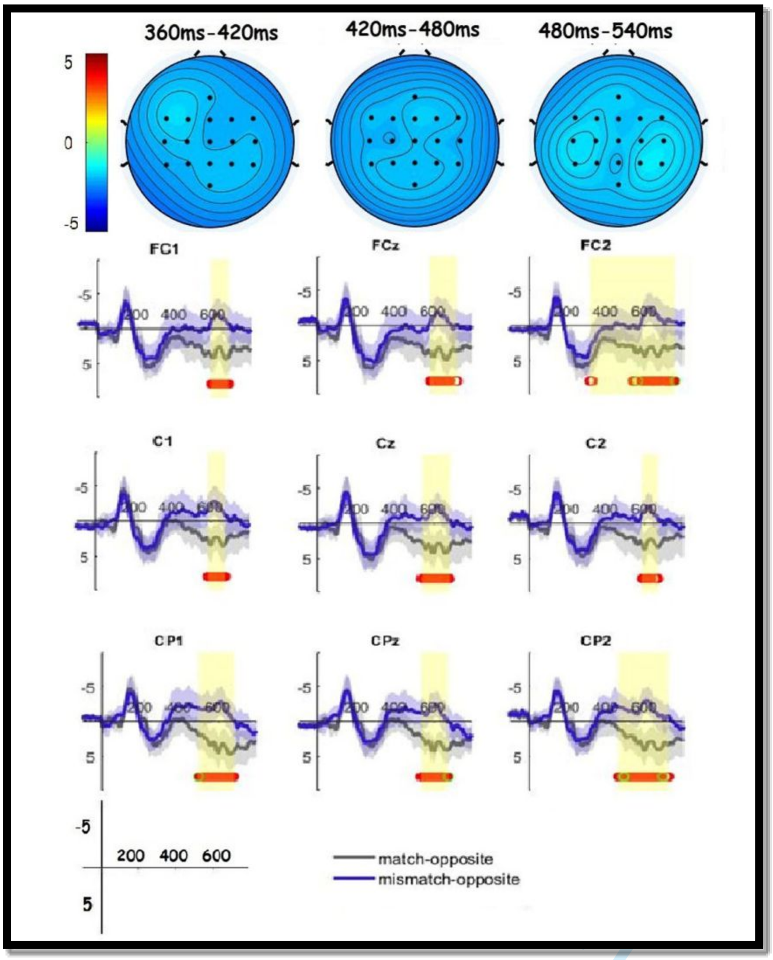

360ms-420ms
420ms-480ms
480ms-540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
-5
200 400 600
5
match-opposite
mismatch-opposite

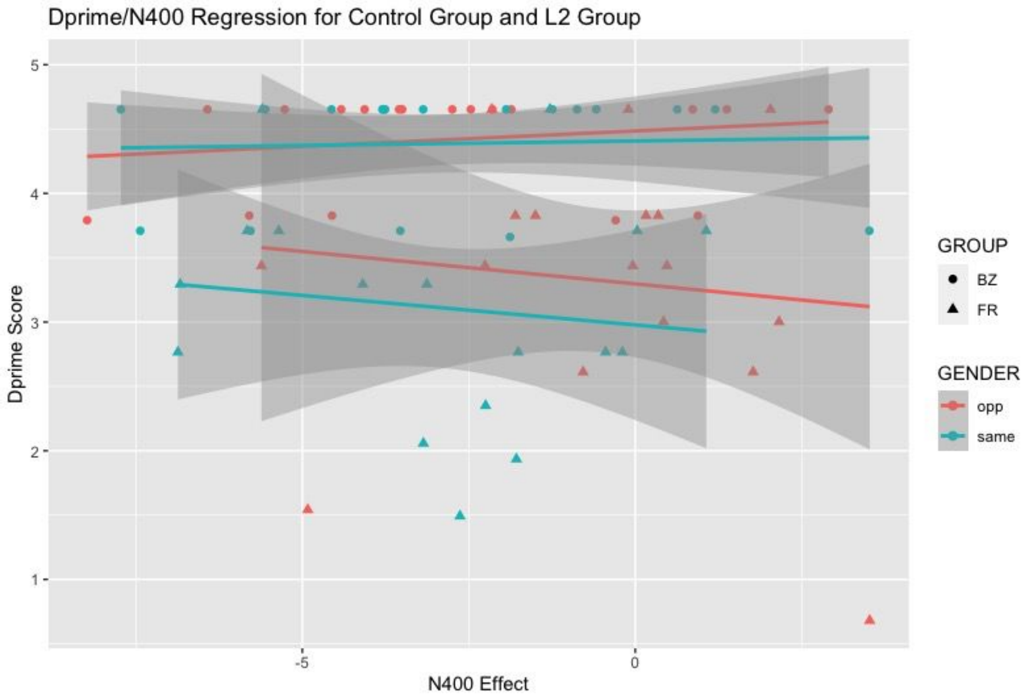


360ms-420ms
420ms-480ms
480ms-540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
-5
5
200
400
600
match-same
match-opposite


360ms-420ms
420ms-480ms
480ms-540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
-5
200 400 600
5
mismatch-same
mismatch-opposite


360ms-420ms
420ms-480ms
480ms-540ms
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
200 400 600
match-same gender
mismatch-same gender


360ms-420ms
420ms-480ms
480ms-540ms
5
0
-5
FC1
FCz
FC2
C1
Cz
C2
CP1
CPz
CP2
-5
200 400 600
5
match-opposite
mismatch-opposite



Dprime/N400 Regression for Control Group and L2 Group
Dprime Score
5
4
3
2
1
-5
0
N400 Effect
GROUP
BZ
FR
GENDER
opp
same

Cross-linguistic gender congruency effects during lexical access in novice L2 learners: Evidence from ERPs

Ana Zappa, Daniel Mestre, Jean-Marie Pergandi, Deirdre Bolger & Cheryl Frenck-Mestre

Herein we present electrophysiological evidence of extremely rapid learning of new labels in an L2 (Brazilian Portuguese) for existing concepts, via computerized games. However, the effect was largely constrained by cross-linguistic grammatical gender congruency. We recorded ERPs both prior to exposure with the second language and following a 4-day training session. Results showed rapid changes in cortical activity, associated with learning. Prior to exposure, no modulation of the N400 component was found as a function of the correct match vs. mismatch of audio presentation of words and their associated images. Post-training, a large N400 effect was found for mismatch trials compared to correctly matched audio-visual trials. However, for learners these results were only obtained for trials on which the L2 words shared grammatical gender in the learners L1 (French). For trials on which the L2 words had the opposite gender in French, no N400 mismatch effect was found post-training. In contrast, behavioral results showed that all L2 words were learned equally as well, independent of gender congruency across Portuguese and French. For control participants who were native speakers of Portuguese, a clear N400 effect was found for mismatch compared to match trials, which was independent of gender congruency. The results demonstrate that grammatical gender overlap in the L1 and L2 influences lexical activation during the initial stages of establishing a new L2 lexicon.

In Brazilian Portuguese, a mouse, no matter which biological sex, is grammatically masculine (*o*$_{\text{masc}}$ *camundongo*$_{\text{masc}}$) whereas a cockroach is grammatically feminine (*a*$_{\text{fem}}$ *barata*$_{\text{fem}}$). The opposite is true in standard French, with grammatically feminine mice (*la*$_{\text{fem}}$ *souris*$_{\text{fem}}$) and masculine cockroaches (*le*$_{\text{masc}}$ *cafard*$_{\text{masc}}$). This arbitrary assignment of grammatical gender is even more apparent for inanimate objects, with opposite gender assignment for trash cans, brooms and chalk across Portuguese and French, despite both languages being derived from Latin. The present study examined how cross-linguistic gender congruency, i.e. the overlap in grammatical gender for nouns across languages, might affect both the acquisition and online processing of a second language (L2) in novice adult learners. Although numerous online studies have provided evidence that speakers of gendered languages are sensitive to gender congruency across languages, during both L2 comprehension and production (cf. Sá-Leite, Fraga & Comesaña, 2019, for a meta analysis) these studies have almost exclusively examined learners who had extensive experience with the L2. We propose a novel approach to this question by starting from the initial stages of exposure to the L2. In addition, we provide both behavioral and electrophysiological measures of performance, which indeed revealed different patterns of the effect of gender congruency.

Numerous psycholinguistic studies have examined the effect of gender congruency (GCE), both within languages in monolinguals and across languages in bilinguals. Monolingual studies have been conducted in the framework of speech production models, which generally assume that grammatical gender is represented independently from other levels of lexical representation, i.e. phonological and semantic, but differ as concerns when and how gender is retrieved (Caramazza, 1997; Cubelli et al. 2005; Foucart et al., 2010; Levelt et al., 1999; Schiller & Caramazza, 2006). Several bilingual studies, discussed below, have also been conducted within this framework. As concerns comprehension, bilingual studies that have examined cross-linguistic GCE have looked at both the interactive nature of bilingual lexical access (Morales et al., 2016; Paolieri et al., 2020) and late bilinguals' ability to use grammatical gender to predict upcoming elements (Hopp & Lemmerth, 2016; Lemmerth & Hopp, 2019). The latter have addressed processing at the lexical level. However, a handful of comprehension studies have also investigated the influence of cross-linguistic gender congruency on syntactic processing (Foucart & Frenck-Mestre, 2011, 2012; Sabourin et al., 2006). We shall address these topics in turn.

Cross-linguistic GCEs have been examined at the lexical level during production, in naming (Bordag, 2004; Bordag & Pechmann, 2007; Costa et al., 2003; Lemhöfer et al., 2008; Morales et al., 2011; Paolieri et al., 2010) and translation (Bordag & Pechmann, 2008; Paolieri et al., 2010; Salamoura & Williams, 2007), as well as in comprehension (Lemhöfer et al., 2008; Paolieri et al., 2020). The pattern of results across studies is both complex and inconsistent. In two independent experiments with German L1-Dutch L2 late bilinguals, Lemhöfer et al. (2008) examined the effects of cognate status and gender congruency on lexical decision times and naming latencies in the L2. They reported robust effects of both factors in both tasks, with no interaction effects. Participants showed faster lexical decision times and naming latencies for L2 Dutch words that shared gender in German and for cognates (cf. Sá-Leite et al., 2019, as well as Costa et al., 2000, for a discussion of cognate effects). In addition, in the naming task, no effect of syntactic structure was found, such that GCE were reported independent of whether participants produced determiner phrases or bare nouns. Similar results were reported by Bordag & Pechman (2007) for relatively inexperienced Czech L1-German L2 late learners, who found GCE in 2 experiments independent of whether participants named bare nouns or nouns preceded by gender-marked adjectives in the L2. In like fashion, Paolieri and colleagues reported faster naming latencies in the L2 (Spanish) for line drawings that shared the same gender in the bilingual participants' L1 (Italian) (eg "falda" and "gonna" *skirt*) compared to those that had opposite gender (e.g., "mesa" and "tavolo" *table*) and this was true for both bare nouns and determiner noun phrases (Paolieri et al., 2010). Paolieri et al. (2010) replicated their results for bilingual participants across 2 experiments; however, in their second experiment both the effect of gender congruency and the critical interaction with Group (bilingual vs. monolingual) were only reliable by participants, suggesting that the GCE was not restricted to the bilingual group and may have been partially driven by characteristics of the items unrelated to gender congruency. Klassen (2016) examined GCE in naming across Spanish and German, which differ concerning the number of gender categories (2 vs. 3), the phonetic regularity of gender marking (overt and highly consistent in Spanish vs. not phonetically salient in German) and complexity (case interacts with gender in German but not Spanish). Results showed facilitated processing for gender congruent nouns compared to incongruent nouns, for both bare nouns and determiner phrases in the group of L1 Spanish-L2 German learners but not L1 German controls. Yet, this was true for nouns that were either masculine or feminine in German but not for neuter gender, which, although indeed incongruent with either gender in Spanish led to faster naming latencies. Hence, the GCE also appears to be constrained by the overlap of gender categories across languages. All of these studies argue in favor of an interactive activation model of processing (Dell, 1986) in which both the L1 gender and phonological form of lexical entries influence L2 processing, and according to which grammatical gender is not stored as an independent feature at the lemma level (i.e. Levelt et al., 1999).

In contrast, Costa and colleagues failed to find GCE across 5 independent experiments in which participants produced NPs in their L2, even when with gender-marked determiners (Costa et al., 2003). This was true independent of whether the bilinguals' two languages had similar gender systems. Costa and colleagues (2003) argued that while semantic representations are shared across languages and commonly activated by lexical entries of either language, the specific grammatical features of a lexical entry, such as its gender, are inherent properties of that entry. Hence, these features would not be shared across languages. Costa et al. (2003) noted nonetheless that they tested highly proficient bilinguals and suggested that less proficient L2 speakers might show greater interaction between their two gender systems, as was indeed reported by subsequent studies (Bordag & Pechmann, 2007; Lemhöfer et al., 2008; Paolieri et al., 2010). In addition, as highlighted by Sá-Leite et al. (2009), other mitigating factors that were not considered by Costa et al. (2003) may have played a role and could explain why this study seems to be the odd man out as concerns finding GCE.

In translation tasks, the results are also mixed. At one extreme, Bordag & Pechmann (2008) reported no cross-linguistic GCE across three translation tasks with Czech-German participants who translated either bare nouns or adjective-noun phrases into the L2. It is of particular interest that the absence of a GCE was reported in translation for the same materials and participant population that produced a robust GCE in production (Bordag & Pechmann, 2007). Bordag and Pechmann (2008) account for the discrepancy in results across tasks as concerns GCE in terms of time course. During picture naming, activation would spread from the concept to the L1 and L2 lemmas in parallel, hence leading to the simultaneous activation of L1 and L2 gender nodes and competition when the two do not match. In translation, the time course would be shifted, such that L1 word forms would activate their lemmas which in turn would activate the L2 lemma and word form. As such, the L2 gender node would only be activated subsequent to the L1 gender node and no direct competition would arise. Salamoura & Williams (2007) reported a different pattern of results, whereby proficient Greek-German bilinguals showed cross-linguistic GCE when they translated gender-marked adjectives along with the noun, but not for bare nouns. The authors argued that gender retrieval occurs only when gender concord must be computed, i.e. within the determiner phrase, in line with certain monolingual models of production (Caramazza, 1997, but see Cubelli et al., 2005). Last, Paolieri et al. (2010) reported GCE during a translation task with advanced Italian-Spanish bilinguals, independent of whether participants produced the bare noun or a determiner phrase. Based on this pattern of results, the authors proposed that lexical selection necessarily entails the activation of gender, in parallel. Lexical items that share gender across languages would enjoy a higher level of activation, hence facilitating both naming and translation. The three competing sets of results and theoretical accounts leave ample room for discussion.

The effect of gender congruency across languages has also been examined during online comprehension. Several studies have approached this topic in the framework of whether bilinguals can use grammatical gender to predict upcoming elements in their L2 (Hopp & Lemmerth, 2016; Lemmerth & Hopp, 2019; Morales et al., 2016). The results from two visual world paradigm studies with Russian-German bilinguals, which used the same design and materials, failed to produce statistically conclusive evidence that gender congruency plays a significant role in the ability of either adults (Hopp & Lemmerth, 2016) or children (Lemmerth & Hopp, 2019) to process gender online in the L2. Another visual world paradigm study, conducted with proficient Italian-Spanish bilingual adults, showed interference from a distracter image when its gender was incongruent with the L1 equivalent, suggesting gender-induced competition (Morales et al., 2016). However, the effects were not significant until after the onset of the target noun, suggesting that co-activation of gender across languages during comprehension may not occur until a certain amount of information has been processed.

Event-related potentials (ERPs) have also been used to measure the effects of cross-linguistic gender congruency on language processing. The most common of these are the N400, the P600 and the LAN, generally elicited by semantic and syntactic violations respectively. The N400, a negative deflection in the waveform with a central-parietal distribution usually observed between 300-500 msec after stimulus onset, is generally thought to reflect semantic integration such that increased N400 amplitude is attributed to processing difficulty that results from attempting to integrate a new element within an existing semantic context (Holcomb, 1993; Kutas & Federmeier, 2011; Kutas & Hilliyard, 1980). The N400 component has also been associated with retrieval/access, with increased N400 amplitude reflecting the effort of retrieving semantic/conceptual information from long-term memory. In this case a reduced N400 is interpreted as signaling facilitated access to lexical information (Delogu et al., 2019). Recently, Bornkessel-Schlesewsky & Schlesewsky (2019) re-examined N400 results in a predictive coding perspective. Accordingly, the N400 (along with other language-related negativities) could reflect precision-weighed prediction errors rather than linguistic processing per se. Another ERP component commonly used in language studies is the P600, which is a positive-deflection in the wave form with a centro-parietal distribution thought to show difficulty in syntactic integration (Kaan et al., 2000; Meltzer & Braun, 2013), as well as the cost of "revising, repairing or reanalyzing an existing (morpho-) syntactic structure" (Delogu et al., 2019, p.2). The P600 is sometimes associated with the Left Anterior Negativity (LAN) (for a debate on the significance of the LAN, cf. Fraga et al., 2021; Molinaro et al., 2015; Tanner, 2015). The P600/LAN has also been associated with conflict/monitoring resolution (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kim & Osterhout, 2005) as well as semantic integration (Brouwer et al., 2017). Finally, the nogo N200 response occurs roughly 200 msec post-stimulus when there is conflict and/or inhibition as concerns processing and the participant's response (Enriquez-Guppert et al., 2010). While not specific to language processing, the N200 can be used to measure response conflict during language processing such as in the study described below.

Rodriguez-Fornells & Münte (2016) recorded ERPs in a Go/Nogo paradigm to probe the effect of grammatical gender congruency (and language switching) across German and Spanish in fluent bilinguals. Compared to monolingual controls, bilinguals showed greater negativity (N200) for incongruent compared to congruent gender trials. While this result may indicate the automatic activation of gender across languages, there are several caveats. First, the task explicitly required participants to retrieve grammatical gender and both languages were actively recruited. Second, participants could potentially predict the incongruent gender trials based on the structure of the experiment. Last, gender incongruence elicited a late ERP component (P600/LPC). Hence, these results do not provide clear evidence for the automatic, early retrieval of gender. Paolieri et al. (2020) also recorded ERPs to examine the effect of gender congruency during the processing of translation equivalents in Spanish and Catalan. Only bare nouns were presented. A small but reliable increase in N400 amplitude was found at central-midline sites for translation pairs that did not share grammatical gender across languages in comparison to those that did, along with increased response times for incongruent pairs, leading the authors to claim that gender is automatically activated during lexical retrieval and elicits competition when different genders are activated across languages. As with Rodriguez-Fornells & Münte (2016), however, participants were actively processing both languages; a stronger demonstration would have consisted in comparing N400 amplitude for bare nouns in the L2 alone as a function of gender congruency. Moreover, no L1 control group was included such that it is not possible to determine whether the effect was driven solely by gender congruency or perhaps by extraneous factors specific to the gender-incongruent pairs. Nevertheless, in a semantic categorization task conducted exclusively in English with Spanish-English late bilinguals and monolingual controls, Boutonnet et al. (2012) reported a late negative ERP component (starting at roughly 400 msec) for trials that did not share gender in Spanish (the L1) with the two preceding items. This effect was specific to the bilingual group. No effects were found for the behavioral measures, in contrast to Paolieri et al. (2020), although the absence of an effect for behavioral measures in the presence of electrophysiological evidence is rather common. The authors argued for the automatic activation of L1 gender in speakers of gendered languages, even when processing exclusively the L2 and in a non-gendered language such as English.

Gender congruency effects have also been examined at the syntactic level during sentence processing, using ERPs. Several monolingual studies have shown that gender agreement violations in sentential context systematically elicit the P600 component (Alencar de Resende et al., 2019; Barber & Carreiras, 2005; Beatty-Martinez et al., 2021; Hagoort, 2003; Frenck-Mestre, 2005; Fraga et al., 2021; Gunter et al., 2000; Popov & Bastiaansen, 2018; Popov et al., 2020; Wicha et al., 2004) and can also produce a LAN (Alencar de Resende et al., 2019; Barber & Carreiras, 2005; Beatty-Martinez et al., 2021; Fraga et al., 2021; Gunter et al., 2000; Popov et al., 2020). The electrophysiological signature of gender concord during sentence processing is nonetheless modulated by various factors, both linguistic (Alencar de Resende et al., 2019; Beatty-Martinez et al., 2021) and task-related (Schacht et al., 2014). In Brazilian Portuguese, Alencar de Resende and colleagues (2019) compared ERP signatures elicited by gender concord violations within the determiner phrase, either between the determiner and noun or the adjective and noun, for nouns with either regular or irregular gender assignment in Brazilian Portuguese. Results showed a biphasic LAN/P600 in response to concord violations, independent of the regularity of gender assignment. However, the amplitude of the P600 evoked by concord violations was greater for regular nouns. The authors suggested that both regular and irregular forms are stored and accessed in similar fashion, i.e. via a single lexical route, but that repair processes are facilitated for regular forms. Beatty-Martinez et al. (2021) reported differences as a function of gender category in Spanish, whereby determiner-noun violations in sentence contexts elicited a biphasic LAN/P600 response for masculine nouns in contrast to the same violations for feminine nouns, which did not elicit a LAN and showed a wider P600 distribution in comparison to masculine nouns. This difference in the electrophysiological response was interpreted as being linked to systematic differences in the representation of the two gender categories.

The P600 is also elicited during L2 sentential processing of gender concord violations for L2 learners whose L1 has grammatical gender (Foucart & Frenck-Mestre, 2011; Sabourin et al., 2006) but also for those who do not (Foucart & Frenck-Mestre, 2012; Dowens et al., 2011; Morgan-Short et al., 2010; Tokowicz & MacWhinney, 2005). Moreover, P600 amplitude for gender concord violations in an L2 is contingent on proficiency and age of acquisition (Nichols & Joanisse, 2016). ERP studies that focused on cross-language gender congruency have provided evidence that the overlap of both syntactic rules, as concerns gender concord (Foucart & Frenck-Mestre, 2011;

Sabourin & Stowe, 2008), and lexical gender across languages (Foucart & Frenck-Mestre, 2011) affect whether gender concord violations in the L2 elicit an electrophysiological response, the type of response (Foucart & Frenck-Mestre, 2012) and its magnitude.

The body of studies cited above has examined L2 gender processing and gender congruency in participants who had several years of experience with and exposure to the L2. Various authors have used learning paradigms with either an artificial language or miniature versions of natural languages to explore how different factors affect gender acquisition after short training periods. Arnon and Ramscar (2012) used an artificial language to test whether acquiring the gender and new lexical labels of known concepts was affected by the sequence of explicit training. Participants who first learned novel words preceded by their gender-marked article within an auditory sentential context, followed by paired associate learning with bare nouns had better learning outcomes, both for gender assignment and noun labels, than those who learned in the opposite order. The authors argued that learning new lexical labels for concepts first via bare nouns blocked the later acquisition of gender assignment in sentential context due to the redundancy of gender in relation to meaning. Brooks and Kempe (2013) examined the implicit learning of nominal gender agreement and case marking in a subset of Russian following a 6-day training session. Results showed that while learners relied on metalinguistic knowledge to acquire case marking, for gender concord they relied on the consistent and transparent morphological cues (feminine being systematically indicated by a final vowel on the noun and agreeing adjective, and masculine by a final consonant) and knowledge of nominal morphology in another known L2. Indeed, the best predictor of acquiring Russian gender agreement was whether the learners had already acquired a Latin language with the same rule for feminine gender. Morgan-

Short and colleagues (2010) compared the processing of gender concord in early learning stages, after implicit versus explicit training in an artificial language. During early stages of acquisition, the ERP signature to gender concord violations differed according to the type of training. Notwithstanding, both implicit and explicit learning groups ultimately attained similar levels of proficiency and exhibited similar patterns of cortical response to gender concord violations at the final stage of acquisition.

To our knowledge, no studies have measured the effect of cross-language gender congruency during the early stages of L2 lexical acquisition in a natural language. The current study aimed to fill this void by examining how cross-linguistic gender congruency might influence processing in an L2 from the very initial stages of learning. We examined French L1 speakers' acquisition of Brazilian Portuguese via interactive computer games, in which L2 Portuguese was presented aurally in full sentences and segmented format, in which grammatical gender within determiner phrases was taught implicitly. Both French and Portuguese have two classes of grammatical gender (masculine and feminine) and require gender concord within the determiner phrase. Whereas French uses the singular definite article le[lə]$_{masc}$ to mark the masculine gender and Portuguese uses o$_{masc}$ (realized as /o/ or /u/), in both languages the singular feminine definite article carries the final phoneme [a] (French: la$_{fem}$ [la], Portuguese a$_{fem}$ [a]). In addition, in Portuguese, the vowel of the definite determiner is generally consistent with the final vowel of the noun (e.g.,"a faca" "the knife" and "o garfo" the fork). It is therefore probable that, even without formal instruction concerning the gender of the Portuguese nouns or the determiner system, French native speakers are able to extract this information from the phonological word forms (Brooks & Kempe, 2013; Denhovska & Seratrice, 2017). Indeed, following the seminal work by Karmilloff-Smith (1979), various experimental studies have shown that native French speakers are sensitive to the regularities present in noun endings and reliably use them as early as age 3 to process and/or assign grammatical gender (cf. Seigneuric et al., 2007 for a discussion and further work with children, and Pérez-Pereira (1991) for similar work in Spanish). In Brazilian Portuguese, incongruent grammatical gender marking within the determiner phrase between adjacent elements (notably the determiner and noun) affects children's ability to process gender assignment as young as age 2 (Corréa & Name, 2003), in support of the hypothesis that young children are sensitive to grammatical information conveyed by the determiner and process gender concord in the DP (cf. Corrêa et al., 2011 for further work). Thus, native speakers of gendered languages show rapid acquisition of the phonological features associated with grammatical gender categories, when present, and the syntactic reflexes of elements that are controlled by the noun. Our French learners undoubtedly exploited these capacities when acquiring the L2 Brazilian Portuguese vocabulary.

We created 4 computer games to teach French native speakers a small lexicon in Brazilian Portuguese. All auditory materials were presented exclusively in Brazilian Portuguese. The games involved both full sentences and individual lexical items, which comprised noun phrases (definite determiner and noun) and verbs. We manipulated gender congruency such that the nouns were either gender congruent or incongruent across the learners' L1 and L2 (cf. Table 1). No instruction was provided concerning grammatical gender, however; nouns were always preceded by the singular definite determiner, i.e. the overtly marked vowel (e.g., a_{fem} $faca_{fem}$ [the knife], o_{masc} $garfo_{masc}$ [the fork]). Participants with no prior knowledge of (Brazilian) Portuguese took part in a 3-day training program, during which they learned a vocabulary of 36 words overall, divided into 12 sentences comprised of 3 verbs and 4 nouns each day via the games. ERPs were recorded both pre- and post-training in a match/mismatch paradigm in which auditory nouns were paired with visual images that either depicted the noun (match) or another learned noun (mismatch). Thus, we were able to follow progression from zero knowledge to the recognition of newly learned L2 phonological word forms.

We hypothesized that learners should be able to fully acquire the L2 vocabulary. Performance was measured both by their accuracy scores pre- and post-training and, crucially, the change in electrophysiological response pre to post-training. Concerning the latter, we expected variation in the N400 component, whereby prior to training match and mismatch trials should not differ in the N400 response, but post-training mismatch trials should evoke an increased N400 response compared to match trials due to difficulties in lexical processing. In relation to the congruency of grammatical gender across the L1 and L2, we hypothesized that it should not affect the ERP response pre-training. In contrast, at the post-training session, it should affect the size of the N400 effect if indeed grammatical gender is automatically activated for speakers of gendered languages (Boutonnet et al., 2012; Dahan et al., 2000; Lew-Williams & Fernald, 2007) and if inhibition results from gender inconsistency across languages (Morales et al., 2016; Rodriguez-Fornells & Münte, 2016).

	L2fem	L2masc
L1fem	a saia (la jupe)	o cachimbo (la pipe)
L1masc	a vassoura (le balai)	o casaco (le manteau)

Table 1. Examples of gender congruency/incrongruency across Brazilian portuguese and French

Methods

Participants. Eighteen right-handed French native speakers (10 women), enrolled as undergraduate students at Aix-Marseille Université, aged 20 to 26 years old (M = 22.8 years, *SD* = 2.4), who had no knowledge of Brazilian Portuguese, and 18 right-handed native speakers of Brazilian Portuguese (16 women) aged 22 to 28 (M = 25.3, SD = 3.4) who were enrolled at AMU in a one year abroad program were recruited for the study. One French participant's data was excluded due to displaying knowledge of the L2 vocabulary (an N400 effect for mismatched pairs) prior to training. All French participants had learned English as a second language throughout secondary school; eight had also learned Spanish as a third language while seven had learned German. None considered themselves fluent in Spanish. The Brazilian participants had all learned English in secondary school; 12 considered themselves novice in French and 6 intermediate. Participants had no history of neurological insult and received monetary compensation in exchange for their participation. All participants gave their written informed consent prior to the experiment and were debriefed about its purpose at its end. The study was approved by the local university ethics committee.

Materials. Thirty-six concrete nouns and 9 transitive verbs were presented orally in Brazilian Portuguese (BP), in sentences and in isolation, across 3 training sessions, with 3 verbs and 12 nouns learned in each session. All materials are presented in the appendix. The items were selected based on their ease of imageability, cognate status with the learners' L1 (French) as well as other Latin languages commonly learned in France (Italian, Spanish) and English, and congruency of grammatical gender across Portuguese and French. Half of the nouns had the same gender across French and Brazilian Portuguese ([a_f panela$_f$] / *la*$_f$ *casserole*$_f$ *"the pot"*) and the other half had the opposite gender ([a_f faca$_f$] / *le*$_m$ *couteau*$_m$ *"the knife"*). Of the 36 nouns, 32 were phonetically marked for gender (16 ending with /a/ and 16 ending with /o/) while 4 were phonetically opaque (2 feminine and 2 masculine nouns). As an added precaution, the overlap in gender between Portuguese and Spanish was also checked for the set of 36 nouns. For the 18 nouns that had the same gender in French and Portuguese, 14 also matched between Portuguese and Spanish and 14 between French and Spanish. For the 18 nouns that had opposite gender across French and Portuguese, 7 had opposite gender between Portuguese and Spanish and 11 had opposite gender between French and Spanish. Nouns were systematically preceded by the definite determiner ($/a/_f$ or $/o/_m$). Nouns that shared gender (SG) and those that had opposite gender (OG) across French and Portuguese were equated across numerous lexical variables: printed mean frequency per million in French (M = 19.71, sd = 21.19) vs. (M = 18.04, sd = 16.05) for SG and OG, respectively (New, Pallier Brysbaert & Ferrand, 2004), mean number of letters SG (M = 6.17, sd =1.70) vs. OG (M = 5.53, sd = 1.62), mean number of phonemes SG (M = 5.50, sd = 1.62) vs. OG (M = 5.47, sd = 1.51) and mean number of syllables SG (M = 2.58, sd = 0.79) vs. OG (M = 2.47, sd = 0.80), grammatical gender SG (9fem/6masc, tested in the Match/mismatch task) vs. OG (8fem/7masc, tested in the Match/mismatch task) and Levenshtein distance SG (M = 5.5, sd = 1.4) vs. OG (M = 6.0, sd = 1.62).

Each of the 9 verbs was paired with 4 different nouns to create 36 declarative sentences in canonical SVO order preceded by a lead in phrase e.g., Esfregar / Scrub [Ele está esfregando a janela / a lareira com a escova / o trapo] / He is scrubbing the window/fireplace with the brush/rag)). Three additional partially transparent verbs (e.g., “pintar” “peindre” paint) and 12 additional partially transparent nouns (e.g., “esponja”, “éponge”, sponge) were selected to familiarize participants with the games and for EEG training. All auditory materials were recorded by a native Brazilian female speaker at 48 kHz (32-bit float) in a professional sound booth, in a single session. The onset of each word within auditory sentences was detected automatically using SPPAS (Bigi, 2015) and manually verified using PRAAT (Boersma & Weenink, 2018). The materials were spliced into individual syntactic units (pronoun + copula, determiner + noun, lexical verb) and individual sentences using Audacity 2.2.1 software.

A subset of the materials was selected for pre- and post-training in a Match/mismatch task (see appendix). Thirty auditory nouns were presented in a 2x2 factorial design defined by the congruency of the auditory noun and line drawing (match vs. mismatch) and gender congruency across languages (same vs. opposite gender). Auditory nouns were paired with 30 line drawings selected from the Snodgrass & Vanderwart (1980) and Alario and Ferrand (1999) standardized image databases (with the exception of two images which were taken from line-drawing internet databases). These images were different from those used for the computerized games (described below), which were selected from internet databases. Each auditory noun was presented twice, once paired with the correct line drawing (match) and once with a line drawing that corresponded to another (to be) learned noun (mismatch). For each Condition (March vs. Mismatch) half of the trials had the same grammatical gender across Portuguese and French and half had opposite gender: Gender (Same vs. Opposite). Hence, a full factorial design was used: Gender (Same vs. Opposite) x Condition (Match vs. Mismatch). For Match trials, where the presented image matched the auditory word, only the gender of the auditory word was in play, which was either the same in Portuguese and French (e.g., *a janlea / la fenêtre* [the window]) or opposite across languages (e.g., *o garfo / la fourchette* [the fork]). For mismatch trials, where the image *did not* correspond to the auditory word, we controlled for the coherency of gender across the distractor image and auditory target word in Portuguese. For example, the auditory word *o cachimbo* (*la pipe* [the pipe]) was paired with the distractor image of a masculine Portuguese word *o lixo* (*la poubelle* [the trashcan]). This prevented participants from noting a mismatch based on the determiner alone if they accessed the name of the distractor image prior to the auditory target word. Three pseudo-randomized lists were created for pre-training EEG testing including 30 "match" pairs and 30 "mismatch" pairs, with 15 same gender and 15 opposite gender pairs for each. Three other pseudo-randomized lists were created for post-training EEG testing, which included the same 30 Match and 30 Mismatch pairs and an additional 30 Semantically related and 30 Semantically unrelated pairs (data reported elsewhere). Participants saw different lists at pre- and post-training testing, with complete counter-balancing of lists across participants such that all lists were seen at both pre- and at post-training sessions.

Procedure

Games. Four computerized games were created in collaboration with the Mediterranean Virtual Reality Center (CRVM). All 4 games involved the auditory presentation via headphones of materials in Brazilian Portuguese accompanied by either static line drawings or animated GIF on a flat screen. Participants' responses and playing behavior (mouse clicks, timing) were recorded throughout each game and feedback was provided. Trials that were not completed successfully were repeated at the end of each game. All participants played the 4 games in the same order and were required to successfully complete a given game prior to engaging in the next. This ensured that all participants had acquired the vocabulary presented throughout the games on each day. In all 4 games, participants initiated a trial by clicking on an audio button image to hear an auditory stimulus. In the first, "exposure" game, participants clicked to hear the 12 sentences, one at a time, which were presented simultaneously with an animated GIF of the action and accompanying objects. In the second "segmentation" game, participants clicked to hear a sentence, which was accompanied by the visual presentation of 5 blank squares at the bottom and 3 in the center of the computer screen. Participants clicked on any of the 5 bottom squares to display a static image and hear the audio file corresponding to it (4 nouns and 1 verb were depicted). Participants had to recreate the auditory sentence by clicking and dragging the 3 correct syntactic elements in order (verb, NP1, NP2) to the center of the screen. Upon correct completion, an animated GIF played along with the auditory sentence. In the third, “verb identity” game, participants clicked to hear a verb and saw three different animated GIFs in the center of the screen, the task being to select the animation that corresponded to the audio file. In the final “memory” game, participants clicked to hear a sentence, then had to find the pair of cards, among 8 presented face down on the screen, that matched the auditory sentence by clicking on the cards individually (a card reverted to blank when another card was selected). The vocabulary depicted across the 8 cards involved a single verb and 4 nouns. When the correct pair was selected, the auditory sentence was replayed along with an animated GIF. Across games 2, 3 and 4, participants were allowed to click on a given item (audio button or card) a maximum of 3 times and time-out was 30 seconds on any given trial. They were encouraged to repeat the materials out loud while playing. Participants were not permitted to take notes during training sessions and were asked not to review what they had learned between sessions. All participants successfully completed all games on each of the three training days, as revealed by their recorded scores.

Training. Participants learned how to play the 4 computer games, prior to actual training, using a miniature auditory vocabulary. This initiation to the games took place directly after the first EEG session. During the initiation and subsequent training, participants were comfortably seated in a sound-attenuated room where they played the games on a 15-inch-screen laptop computer while wearing headphones. Training occurred over 4 consecutive days. On each of the first three days, participants were exposed to 3 verbs and 12 nouns, comprised in 12 auditory sentences, via the 4 games, with the vocabulary repeated across the games. Each session lasted roughly 25 minutes, with each game taking 5-10 minutes. The fourth day consisted of a 40-minute review day, where participants played all the games with the entire new lexicon (9 verbs and 36 nouns).

EEG Match-Mismatch task. A trial began with the presentation of a centered fixation cross for 250 msec that was replaced by a centered black and white line drawing for 1 second followed by an auditory word presented over speakers. At the offset of the auditory word, a visual "yes/no" prompt was presented and participants were requested to judge whether the auditory word matched the visual image or not on a button box. A visual blink prompt was presented for 2s following the response. During the pre-training session, 2 blocks of 30 trials, with 15 "match" and 15 "mismatch" pairs in each, were presented, preceded by three warm-up pairs. During the post-training session, an additional 30 trials were presented per session, comprising 15 semantically related and 15 unrelated pairs (data reported elsewhere), for a total of 60 trials per session. Short rest periods were allowed between blocks. Participants were asked to remain still and to blink at the prompt. Behavioral responses to the questions were recorded. The session lasted roughly 30 minutes.

ERP Data acquisition and analysis. Electroencephalographic (EEG) activity was recorded continuously from 64 scalp locations over frontal, temporal, central, posterior temporal, parietal and occipital areas of the left and right hemispheres and midline.

Individual electrodes were adjusted to a stable offset lower than 20μV. EEG data were sampled online at 512 Hz. Blinks and horizontal eye-movements were monitored by means of electrodes placed beneath the left eye and at the outer canthus of the right eye. Electrodes were placed on the left and right mastoids for referencing offline. Periods spanning from -100 pre- stimulus onset to 1100 msec post stimulus onset were used post-recording for analyses. A low pass digital filter of 30 Hz was applied post-recording. Trials contaminated by ocular-motor or muscular artifacts were excluded using automated routines that were manually checked. The percentage of trails retained for analyses was 88% for the same gender condition and 87% for the opposite gender condition.

Results.

To determine the pattern of data, we ran a series of comparisons for both behavioral and electrophysiological measures. One the one hand, we compared the data for the L2 learner group across sessions. In addition, we compared the L2 learners' data post-training to that of the Brazilian control group. When indicated, we ran subsequent independent models (in each session, for the L2 learners and for each group).

Behavioral measures.

Game-playing accuracy

Errors took the shape of time-outs. Accuracy was at ceiling (over 96% across all games) for all participants. The low level of variability during game play in the review session did not warrant further analyses (cf. Table 2a).

Verb game	Segmentation game	Memory game
22/648 (3.4%)	3/864 trials (0.3%)	3/864 trials (0.3%)

Table 2a. *Total number and percentage of time-outs across the different games for all L2 participants*

Accuracy in the EEG match-mismatch task.

The percentages of correct responses are presented in Table 2b and d prime scores are presented in Figure 1 for the L2 learners at the pre- and post-training session, and for the Brazilian control group. The data were modeled using linear mixed effect regressions, with the *LmerTest* package (Kuznetsova & Christensen, 2017) implemented in R (R Core Team, 2017). For both measures, we first modeled the data for L2 learners, pre- and post-training. The model for the percentage of correct responses included the sum coded fixed factors Training session (Pre vs. Post), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions, with random intercepts for Participant and Item and a random slope for Condition:Gender. The model for d prime included the sum coded fixed factors Training session (Pre- vs. Post), Gender (Same vs. Opposite) and their interaction, with random intercept for Participant. We subsequently compared L2 learners to Brazilian controls at the post training session. For accuracy, the model included the sum coded fixed factors Group (Control vs. L2), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions, with random intercepts for Participant and Item. The same model was applied for d prime, without the factor Condition.

Accuracy.

For L2 learners, the first model, comparing pre- and post-training sessions revealed an effect of Session ($\beta = 1.35$, se = 0.07, z = 18.85, $p < .001$), the interaction between Gender and Condition ($\beta = 0.19$, se = 0.08, z = -2.25, $p < .001$) and the interaction between Gender, Condition and Session ($\beta = 0.14$ se = 0.07, z = 2.02, p

< .05). Subsequently, models were run on the L2 data for the pre- and post-training session independently. Pre-training, the treatment coded model, revealed an interaction between Gender and Condition ($\beta = -1.32$ se = 0.28, $z = -4.74$, $p < .001$), and a simple effect of Condition for Opposite gender trials ($\beta = 0.51$, se = 0.19, $z = 2.74$, $p < .01$). The re-leveled treatment coded model showed a simple effect of Condition for Same gender trials ($\beta = -0.81$ se = 0.21, $z = -3.92$, $p < .001$). Pre-training, L2 participants showed a bias to respond positively for Same gender trials and negatively for Opposite gender trials, although accuracy remained at chance levels. Post-training, the sum coded model revealed only an effect of the Intercept for L2 learners ($\beta = 2.78$, se = 0.26, $z = 10.68$, $p < .001$), due to accuracy being higher than chance. No other effects were significant.

Last, the model comparing L2 learners' performance at the post-training session to the native Brazilian control group failed to converge due to the extremely low variability and high level of accuracy.

	Match	Mismatch
L2: Pre-training		
Same Gender	57% (49)	38% (48)
Opposite Gender	36% (48)	47% (50)
L2: Post-training		
Same Gender	92% (27)	92% (29)
Opposite Gender	92% (27)	90% (27)
Brazilian control group		
Same Gender	95% (21)	100% (0)
Opposite Gender	98% (26)	100% (0)

Table 2b. *Percentage of correct answers (standard deviations in parentheses) in the Match-Mismatch task for L2 learners pre- and post- learning, and for Brazilian controls, as a function of Gender (same vs. opposite) and Condition (Match vs. Mismatch)*

D-prime.

The model comparing L2 learners' data pre- and post-training revealed only an effect of Session ($\beta = 1.76$, se = 0.105, $t = 16.90$, $p < .001$). Participants' d prime scores (i.e. their ability to correctly discriminate between Match and Mismatch trials) increased significantly from pre- to post-training and this did not depend on the gender congruency of trials.

The model comparing d prime scores for L2 learners post-training and the native Brazilian control group revealed an effect of Group ($\beta = 0.59$, se = 0.102, $t = 5.80$, $p < .001$), due to Brazilian controls showing higher d prime scores. No other effects were significant.

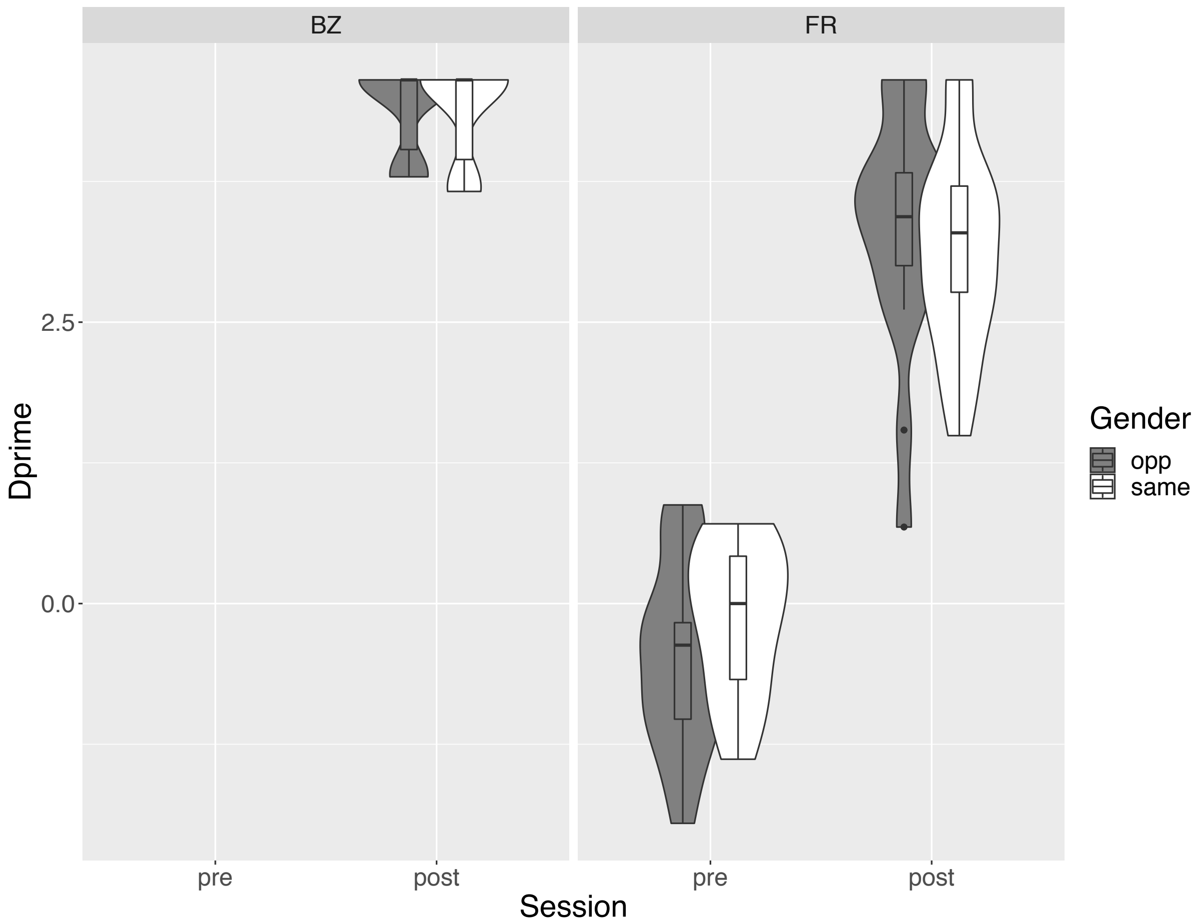


Figure 1. *D prime scores for Brazilian controls and for L2 learners at pre- and post-training sessions*

ERP analysis

The ERP data were modeled using linear mixed effect regressions, with the *LmerTest* package (Kuznetsova & Christensen, 2017) implemented in R (R Core Team, 2017) for the mean voltage amplitudes in the N400 time window, calculated 300-600 msec post onset of the auditory noun for correct trials. This window was based on prior literature and confirmed by permutation tests conducted across the entire epoch (see below). Data were trimmed in R to remove outliers (1% of the data were excluded). Models were performed independently over midline sites (Fz, FCz Cz, CPz, Pz), frontal-central sites (FC1, FC3, FC5, FC2, FC4, FC6, C1, C3, C5, C2, C4, C6) and centro-parietal sites (CP1, CP3, CP5, CP2, CP4, CP6, P1, P3, P5, P2, P4, P6). Below we report the results from the maximal random-effects structure (Barr, et al., 2013).

In addition to the LMER models, to test the hypothesis of a significant difference in the ERP amplitude between conditions, a two-tailed permutation test (1000 random partitions) was carried out over the 1-second post-stimulus time window for all electrodes entered into the models. Statistically significant differences were taken into consideration only if they persisted for 10 msec or more, which corresponds to an interval of 5 samples, given a sampling rate of 512Hz. The results of these tests are visible in Figures 2a through 4b for 9 central electrodes commonly associated with the N400 effect and voltage maps that included a larger array of electrodes.

Pre vs. Post-training: L2 learners

To examine the effect of training, we ran a first model, summarized in Table 3a, which included the sum-coded fixed factors Session (Pre vs. Post), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions, with random intercepts for Participant and Item. Condition included a random slope for Participant and for Item. The model revealed a three-way interaction of Condition:Gender:Session at all electrode sites. The data were modeled independently thereafter for each training session. Pre-training, no effects were found for any factor at any ROI (cf. Table 3b).

Figures 2a and 2b show the mean Match-Mismatch ERP response for Same and Opposite gender conditions, respectively, for illustrative purposes.

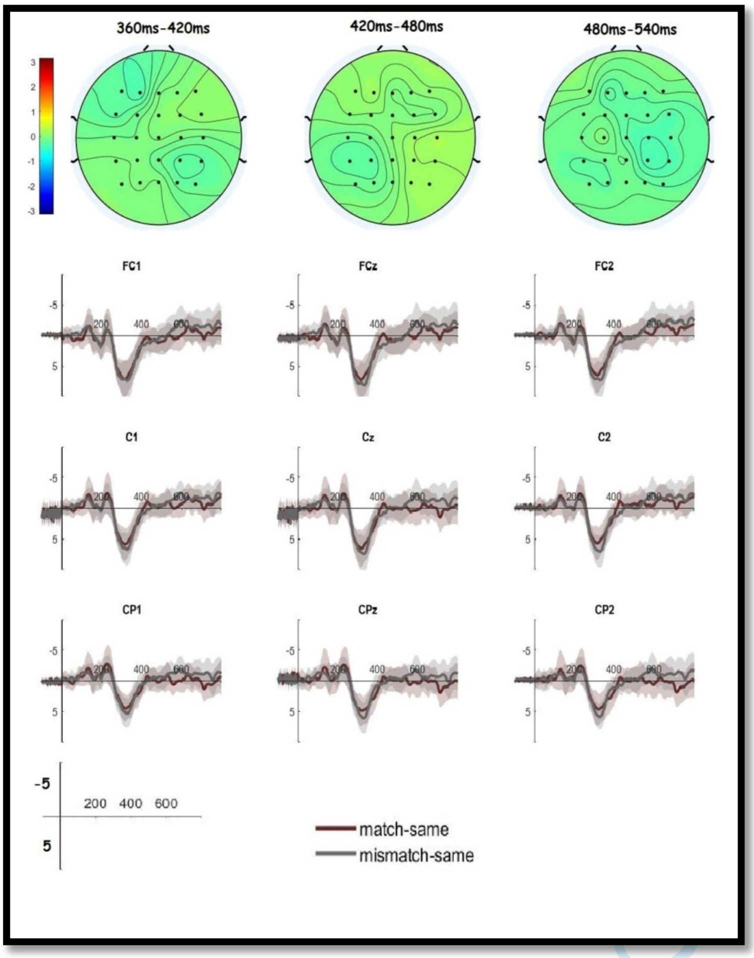


Figure 2a. *Mean voltage ERPs (and SD) for 9 central electrodes in the pre-training session for L2 learners as a function of Condition (Match vs. Mismatch) for nouns with same gender across languages. Permutation tests revealed no significant differences.*

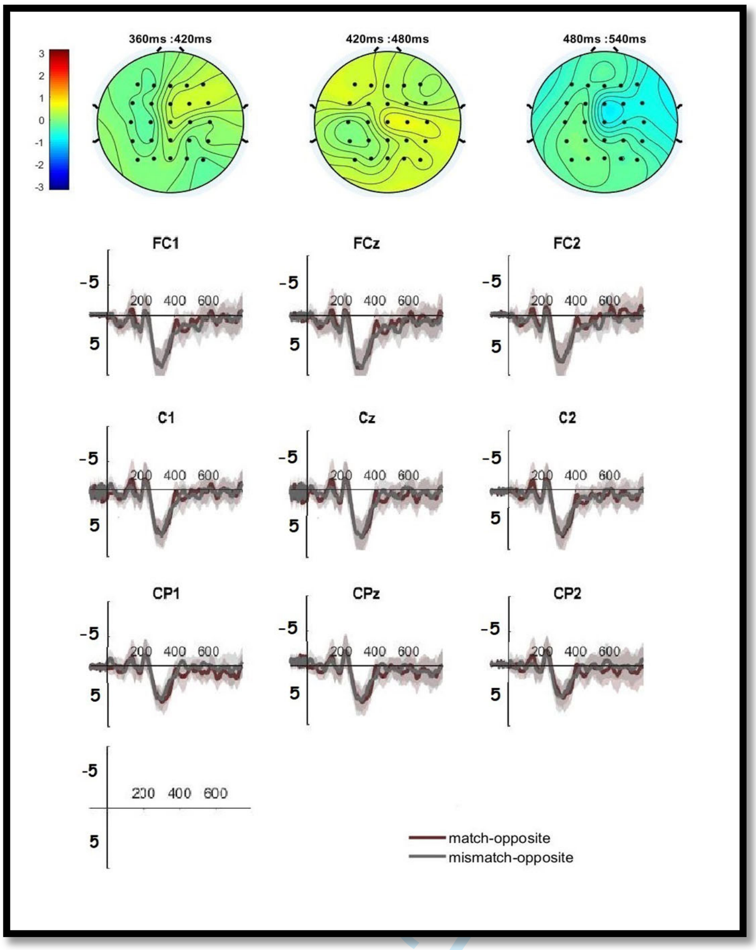


Figure 2b. *Mean voltage ERPs (and SD) for 9 central electrodes in the pre-training session for L2 learners as a function of Condition (Match vs. Mismatch) for nouns with opposite gender across languages. Permutation tests revealed no significant differences.*

Post-training, sum coded models performed independently at all 3 ROI revealed a significant interaction of Condition x Gender at all sites (cf. Table 3c). Models of simple effects (cf. Tables 3d and 3e) revealed a significant effect of Condition for same gender trials at all ROI (midline: β = -3.40, se = 0.770, t = -4.42, $p < .001$; frontal-central: β = -2.60, se = 0.682, t = -3.82, $p < .002$; central-parietal: β = -2.91, se = 0.577, t = -5.05, $p < .001$) but no effect of Condition for opposite gender trials (midline: β = -0.27, se = 0.757, t = -0.36 $p<0.73$; frontal-central: β = -0.52, se = 0.644, t = -0.81,p

< .43; central-parietal: β = -0.59, se = 0.613, t = -0.97, p < .35). These effects are shown in Figures 3a and 3b.

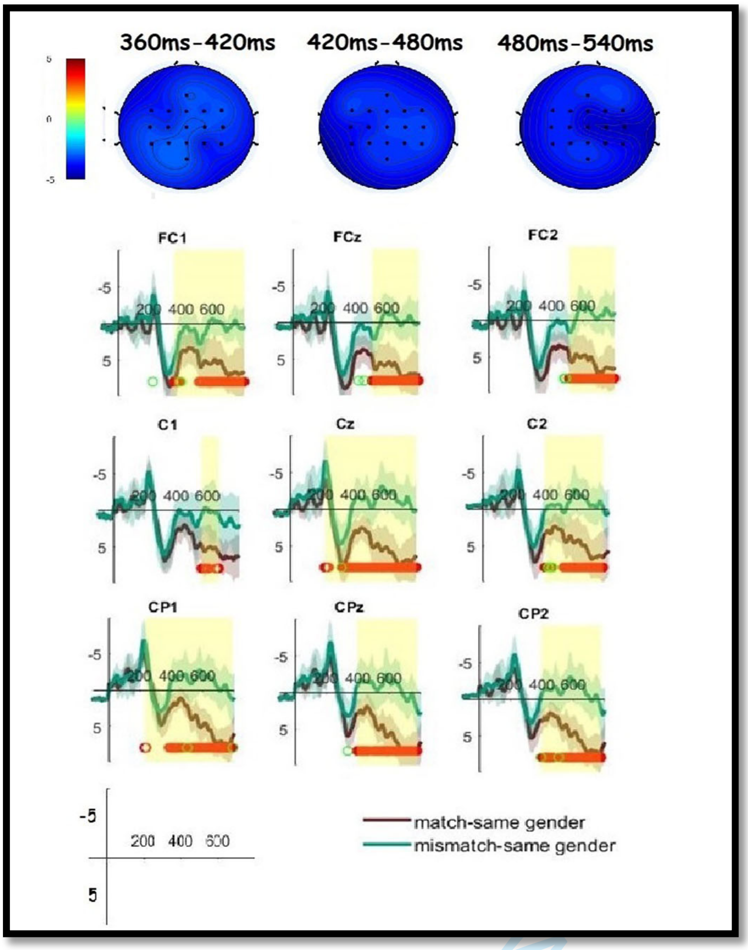


Figure 3a. *Mean voltage ERPs (and SD) for 9 central electrodes in the post-training session for L2 learners as a function of Condition (Match vs. Mismatch) for nouns with same gender across languages. Permutation tests are shown in red (p=.05) and green (p<.05) across the entire epoch.*

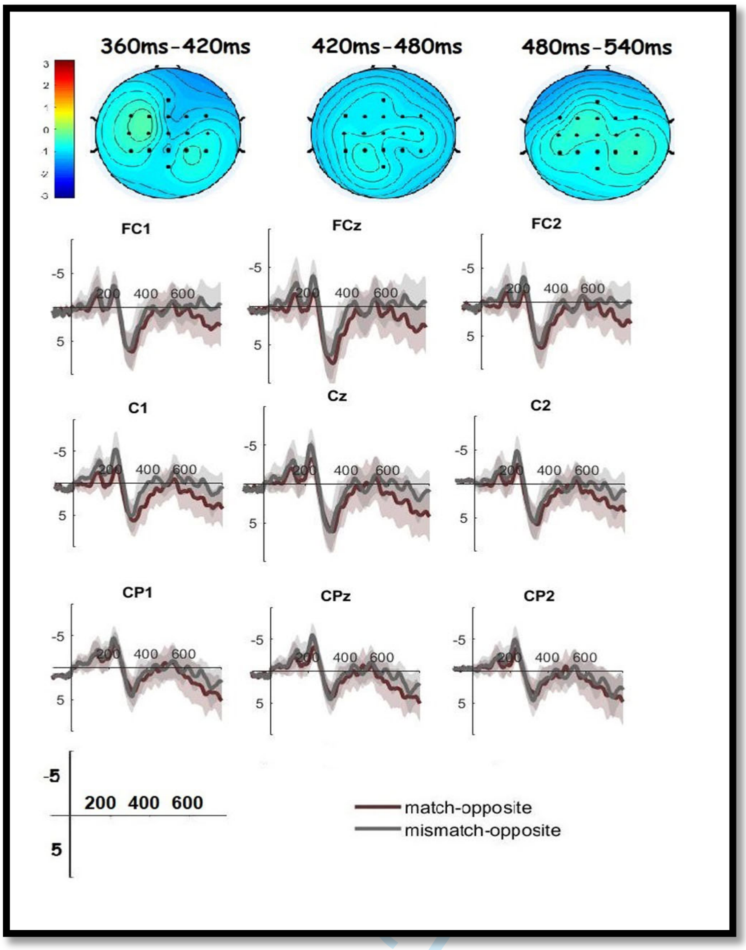


Figure 3b. *Mean voltage ERPs (and SD) in the post-training session for L2 learners for 9 central electrodes as a function of Condition (Match vs. Mismatch) for nouns with opposite gender across languages. Permutation tests revealed no significant effects.*

In addition, for Match trials there was an effect of Gender at all three ROI (cf. Table 3e) (midline: $\beta = 2.08$, se = 0.422, t = 4.92, $p < .001$; frontal-central: $\beta = 1.69$, se = 0.584, t = 2.89, $p < .01$; central-parietal: $\beta = 1.27$, se = 0.257, t = 4.96, $p < .001$), while for Mismatch trials the effect of Gender was only present at midline (midline: $\beta = -0.91$, se = 0.434, t = -2.09, $p < .04$; frontal-central: $\beta = -0.38$, se = 0.805, t = -0.48, $p<0.64$; central-parietal: $\beta = 1.06$, se = 0.89, t = -1.19, $p < .0.25$). These effects are depicted in Figures 3c and 3d.

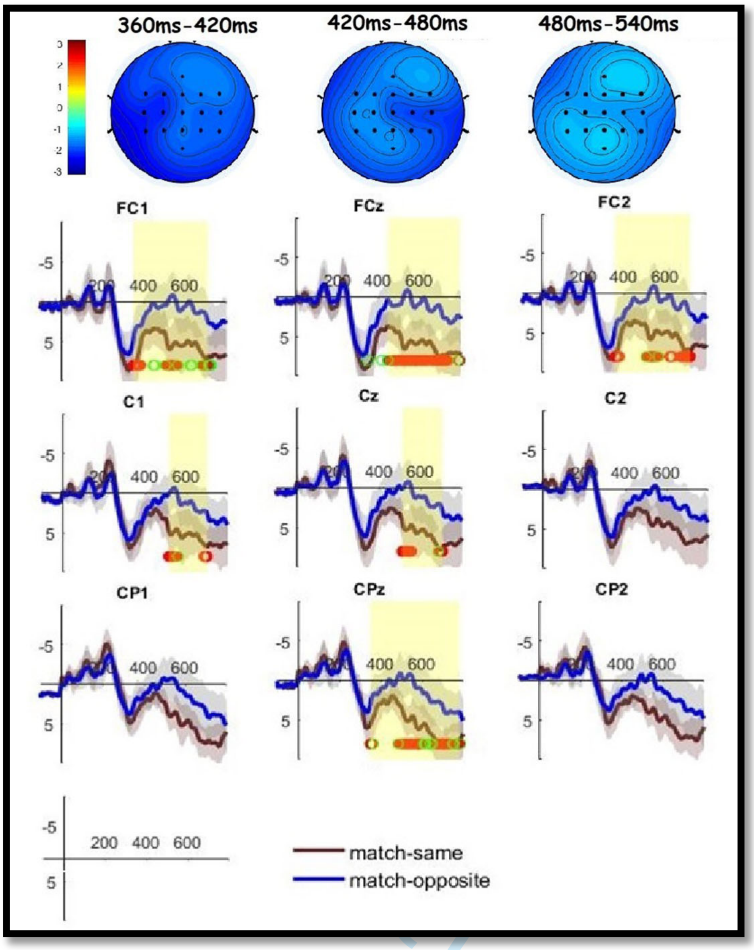


Figure 3c. *Mean voltage ERPs (and SD) for 9 central electrodes in the post-training session for L2 learners as a function of Gender across languages (same vs. opposite) for Match trials. Permutation tests are shown in red (p=.05) and green (p<.05) across the entire epoch.*

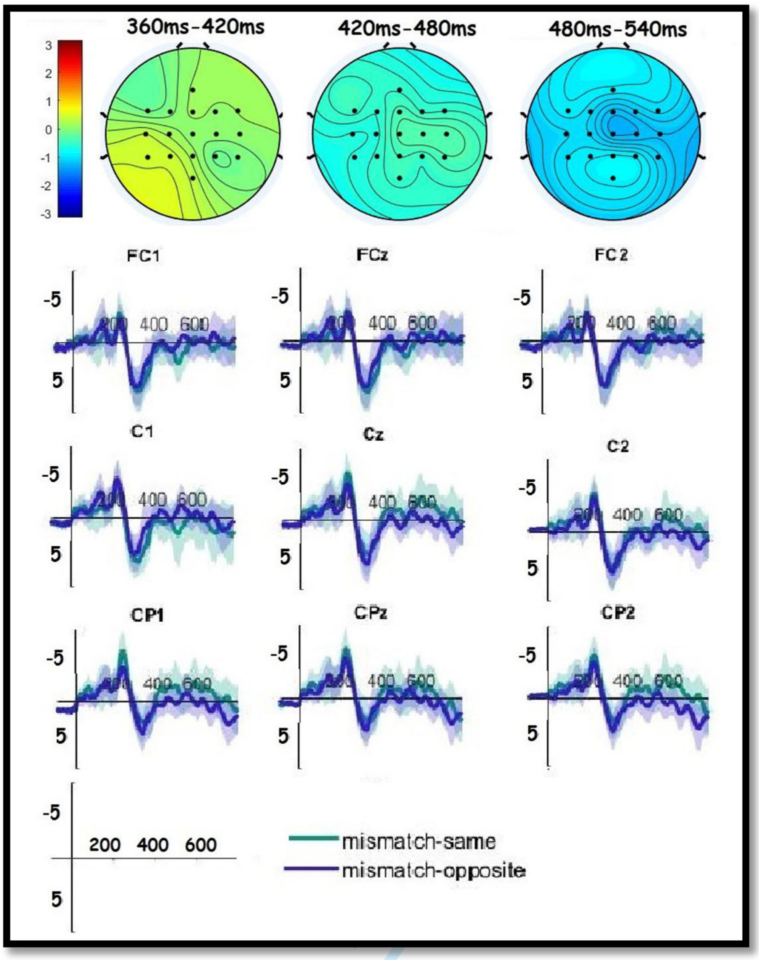


Figure 3d. *Mean voltage ERPs (and SD) for 9 central electrodes in the post-training session for L2 learners as a function of Gender across languages (same vs. opposite) for Mismatch trials. Permutation tests revealed no significant effects.*

Post-training: Native controls vs. L2

The first model, summarized in Table 4a, included the sum-coded factors Group (Native vs. L2), Gender (Same vs. Opposite), Condition (Match vs. Mismatch) and their interactions. Condition included a random slope for Participant and for Item. The model revealed an interaction of Condition:Gender:Group at all sites. For native controls, there was an N400 effect, with Mismatch trials producing a larger N400 amplitude than Match trials, which was independent of Gender congruency across languages. For L2 learners, the N400 effect interacted with Gender congruency. The data were modeled independently thereafter for the Brazilian control group, using the same model structure as above without the fixed factor Group. For native speakers, the effect of Condition was significant at all sites due to greater mean N400 amplitude for mismatch than match trials. At frontal central sites, there was also an effect of Gender, due to a larger N400 for same gender trials. Crucially, Condition did not interact with Gender at any site (see Table 4b and Figures 4a and 4b, for same and opposite gender respectively).

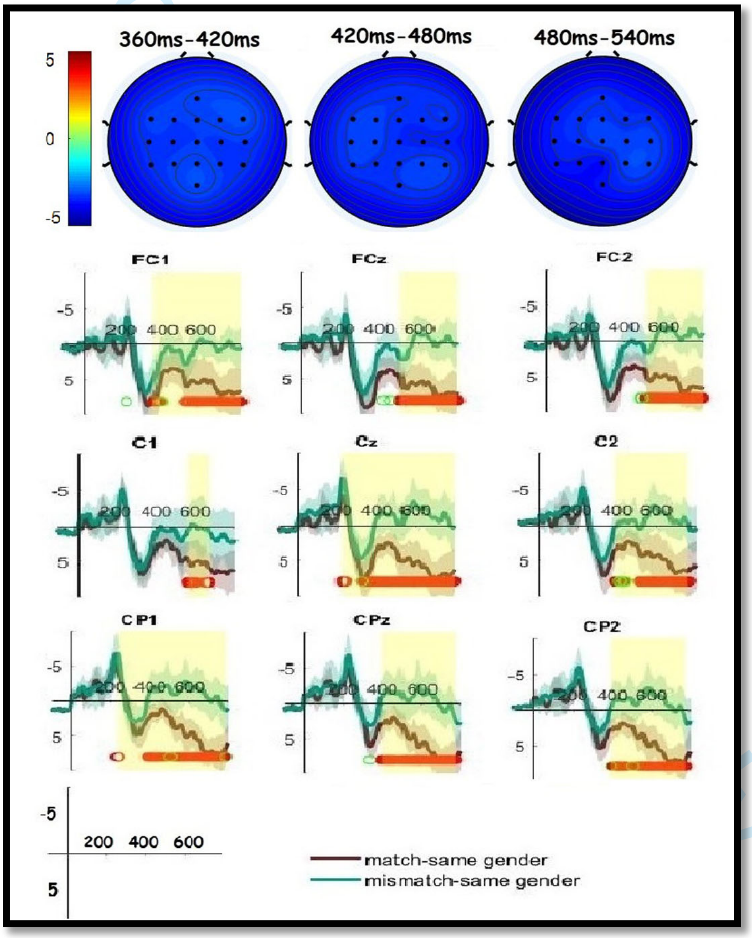


Figure 4a. *Mean voltage ERPs (and SD) for native Brazilians for 9 central electrodes as a function of Condition (Match vs. Mismatch) for nouns with same gender across languages. Permutation tests are indicated in red (p=.05) and green (p<.05) across the entire epoch.*

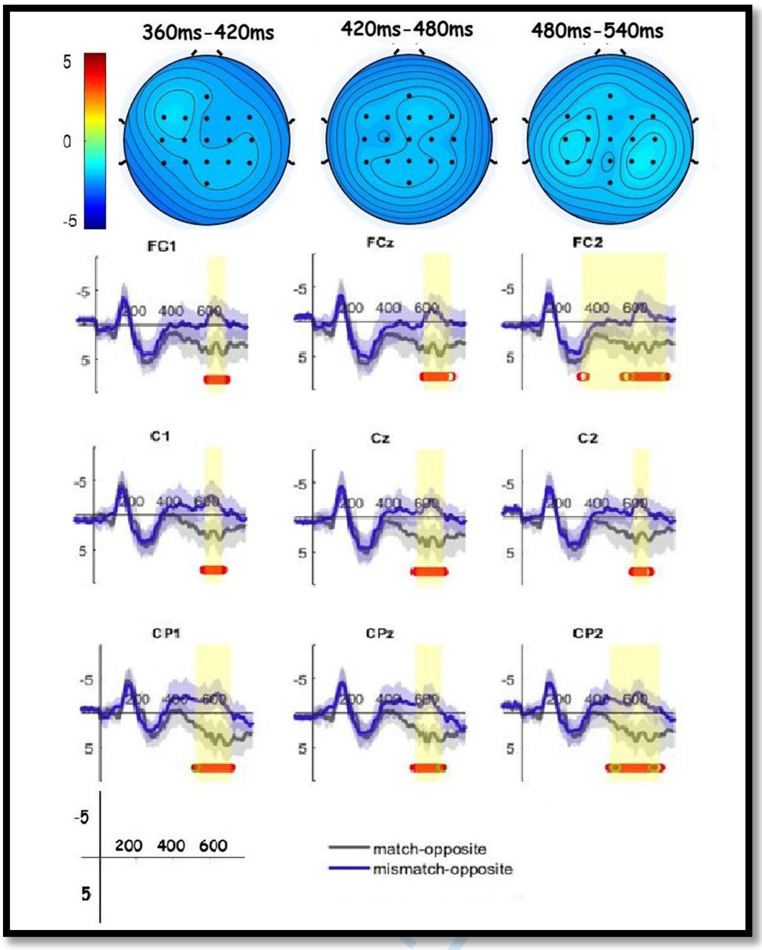


Figure 4b. *Mean voltage ERPs (and SD) for native Brazilians for 9 central electrodes as a function of Condition (Match vs. Mismatch) for nouns with opposite gender across languages. Permutation tests are indicated in red (p=.05) and green (p<.05) across the entire epoch.*

Correlation between D prime score and N400 effect

Pearson's correlations were performed in order to determine whether there was any correlation between the ability to correctly identify match trials (d prime) and the magnitude of the N400 effect (cf. Tanner et al., 2013). We found no correlation between d prime scores and the magnitude of the N400 effect for either the L2 learners (same gender $r(16)=0.07$, p = .72, opposite gender $r(16) = 0.36$, p = .15) or the Brazilian control group (same gender $r(16) = 0.04$, p = .84, opposite gender $r(16) = 0.18$, p = .46). These results are depicted in Figure 5.

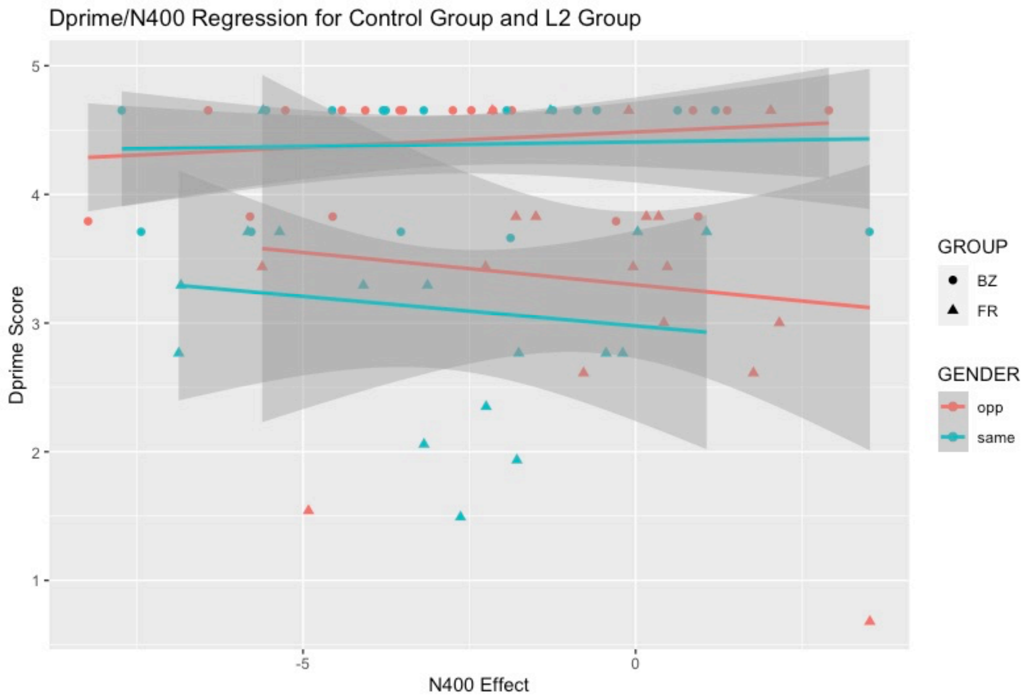


Figure 5. *Regression of N400 effect as a function of D prime score for Brazilian control group and L2 learners*

Discussion

Our study revealed clear cross-linguistic gender congruency effects (GCE), from the earliest stages of acquiring a second language. This was apparent in the electrophysiological trace of lexical access, as measured by the N400. L2 learners demonstrated a clear N400 effect for mismatched visual-auditory pairs post-training, but only for newly learned nouns that shared grammatical gender across their native (French) and newly acquired language (Portuguese). No modulation of the ERP response was found as a function of the match between auditory words and visual stimuli for nouns that had opposite gender across the two languages for these learners. In addition, the effect of gender congruency was visible in the N400 modulation for match trials, for which the N400 response was increased for nouns that had opposite gender across the L2 and French compared to nouns that shared grammatical gender across languages. Hence, cross-linguistic GCE were clearly reflected in the automatic cortical response associated with lexical processing (Kutas & Federmeier, 2011). It is important to note that in the control group of native Brazilian Portuguese speakers, who on average were novice French speakers, only a robust N400 mismatch effect was found, which was independent of gender congruency. Otherwise stated, the cross-linguistic GCE revealed by the N400 was specific to the L2 learners processing nouns in the newly learned language. Importantly, this effect was found following only 4 days of training using interactive computerized games.

Previous studies on the effect of cross-linguistic gender congruency have examined this question in populations that had several years of formal learning of the second language (Bordag, 2004; Bordag & Pechmann, 2007, 2008; Costa et al., 2003; Lemhöfer et al., 2008; Morales et al., 2011; Rodriguez-Fornells & Münte, 2016; Salamoura & Williams, 2007). Our training study allowed us to examine this question from the earliest stages of acquisition. In addition, our design has the distinct advantage of presenting only the newly learned language. In several studies that have shown cross-linguistic GCE, participants had to actively process their native and second language simultaneously due to task requirements (switching between languages, translating, or ignoring embedded L1 words during L2 production; Bordag & Pechmann, 2007, 2008; Costa et al., 2003; Rodriguez-Fornells & Münte, 2016; Salamoura & Williams, 2007). The necessity to maintain both languages active may have played a role in evoking gender congruency effects in these studies. This cannot be claimed for the present results. Indeed, our study did not require L2 participants to overtly produce or consciously activate their L1. Nonetheless, that the L1 lexicon, and more specifically the grammatical features of L1 candidates, became active during L2 processing was readily apparent in the ERP data. Our results are in line with those reported by Boutonnet et al. (2012) who showed modulation of a late negative component as a function of whether triplets of words, presented in English, all shared the same lexical gender in the participants' native language, Spanish. Thus, as in the current study, even in conditions where the L1 was physically absent, it played a significant role in processing. More specifically, native speakers of "gendered" languages automatically and irrepressibly activate the L1 gender of nouns, even when processing words exclusively in the L2.

In many studies, gender congruency effects have been reported both within a given language and across languages when participants were required to produce a definite article prior to the noun (Costa et al., 2003; Salamoura & Williams, 2007); although others have shown GCE when participants produced bare nouns (Bordag & Pechmann, 2007; Cubelli et al., 2005). In light of this, it is of interest to note that, while our participants did not produce the lexical items, all auditory nouns were preceded by the definite determiner (e.g., /oparafuzo/ "the screw" and /aluva/ "the glove"). It is an empirical question whether the gender congruency effects that we obtained would occur in the absence of the determiner. Given that our participants showed evidence that they had acquired the gender of the L2 nouns, it is possible that they may have retrieved this information during processing, either from a stored representation of the noun or from the morphology of the word form itself, which may then have been the source of interference (cf. Gollan and Frost, 2001, for a discussion of different routes to stored gender information). Further work is necessary to determine the locus of the interference we found.

Our results show rapid learning of the L2 vocabulary, as demonstrated by both the establishment of an N400 response to newly learned words and by ceiling level accuracy post-training. Concerning the cortical response, modulations of the N400 have been noted in association with L2 learning and/or artificial languages in several studies.

This has been demonstrated in L2 studies that have used longitudinal designs to investigate changes in cortical activity over the course of learning (Chun, Choi, & Kim, 2012; McLaughlin, Osterhout, & Kim, 2004; Stein et al., 2006; Yum, Midgley, Holcomb & Grainger, 2014). In a seminal study, McLaughlin and colleagues (2004) found that L2 pseudowords elicited a larger N400 than learned L2 words following 14 hours of classroom instruction (McLaughlin et al., 2004). Crucially, these effects were not seen behaviorally; when making overt lexical decisions, learners performed at chance. Hence, L2 learners were sensitive to the prior exposure to word forms, as shown by the N400 effect, even when they could not consciously identify these forms. In contrast, the semantic integration of these newly learned lexical items only occurred following 60 hours of instruction, as indexed by reduced N400 amplitude for L2 target words preceded by semantically related primes. Similarly, Soskey and colleagues (Soskey et al., 2016) reported the gradual instantiation of L2 words, as indexed by modulation of the N400, across a semester of learning. It is important to note, however, that these studies reported cortical changes due to L2 meaning integration following extended L2 training whereas we found that participants accessed the meaning of newly acquired L2 words after only 3 hours of learning.

Artificial language paradigms have been used to ascertain whether novel L2 words can be associated quickly with meaning. Breitenstein and colleagues (Breitenstein et al., 2007) used associative learning, where a spoken word was paired with the image of an object with increasing statistical probability over multiple trials and found that after 5 days of training, newly learned words facilitated (in the form of shorter response latencies) the processing of related pictures, indicating integration into existing lexical networks. A similar magnetoencephalography (MEG) study by Dobel et al. (2010) showed a reduction in the mN400 (the MEG component comparable to the

ERP N400 component) to correct images preceded by trained spoken words from pre- to post-training, indicating that trained words had become associated with existing conceptual representations. Our results corroborate these findings, showing the acquisition of a small L2 vocabulary following three 25-minute training sessions and one 40-minute review session over the course of 4 days, as manifested by the establishment of an N400 response from pre- to post-training and increased accuracy, from chance to ceiling level.

Our study clearly demonstrates that learners were sensitive to the grammatical gender of the newly learned words, despite there having been no formal or explicit instruction concerning gender and even though the stimuli were only presented aurally. This implies that our participants segmented the auditory stimuli into the determiner and noun and inferred gender information from the properties of the speech signal. Segmenting the audio signal into its syntactic elements is notoriously difficult during second language acquisition (Altenberg, 2005), which is why we created a game that specifically segmented the auditory sentences and required participants to recognize the meanings of the different elements and assemble them in the correct order to recreate the auditory sentence. However, while determiner phrases and verbal phrases were explicitly segmented, the determiner phrases were presented as a single unit (e.g., /okaSimbo/ "the pipe" and /asaia/ "the skirt"). Learners could, in theory, have interpreted the determiner phrase as a whole rather than segment it into the determiner and noun, as indeed there is evidence for in young children. For example, in early stages of acquisition French children may produce forms that reveal segmentation errors (e.g., "*le loiseau*" and "*le noiseau*" stemming from the speech signal "*l'oiseau*" and "*un oiseau*" «the bird / a bird» Clark, 2009). Even so, the regularity of the morphological form for the determiner (/a/ or /o/) preceding the noun and the concurring final phoneme of nouns reduces this possibility, as discussed below. Moreover, the majority of participants had formally learned Spanish as a third language throughout secondary school, which may well have prompted them to capitalize on the partial overlap of gender concord rules in Spanish and Portuguese (cf. Brooks & Kempe, 2013).

Various studies with either natural or artificial languages have shown that following implicit training, adult learners rapidly deduce the rules that govern grammatical gender assignment (Öttl & Behne, 2017) and gender concord (Denhovska & Serratrice, 2017; Morgan-Short et al., 2010). Concerning gender assignment, our materials provided a clear phonetic cue, as outlined above, but no semantic information was associated with the gender of nouns. This differs from the materials learned in an artificial language (Öttl & Behne, 2017) in which gender suffixes on the noun were determined by the biological gender of stimuli. Concerning gender concord, Morgan-Short et al. (2010) found no difference in learning as measured by behavioral (d prime) or cortical sensitivity (P600 response) to determiner-noun gender concord violations as a function of the type of training (implicit or explicit) at the end of training. Using a miniature set of Russian nouns and adjectives, in which adult learners were exposed to noun-adjective gender concord in short written sentences, Denhovska and Seratrice (2017) showed that even under implicit learning, where emphasis was placed on learning the meanings of the sentences and no mention was made of the underlying grammatical rules, participants readily acquired these rules. Moreover, no difference in behavioral accuracy was found for grammatical judgments as a function of the type of instruction (implicit or explicit), although only those who received explicit instruction were able to produce the grammatical rules governing gender concord above chance level. The present results are in line with those found in the above studies, showing that participants rapidly acquire grammatical gender concord rules in a newly learned language following short training sessions, even in the absence of formal instruction.

The pattern of results we obtained suggests that gender congruency effects play an early role during lexical access. This question, i.e. whether grammatical gender exerts an early influence on lexical access or only later, during lexical selection, has been examined in various monolingual studies. Eye-tracking experiments have demonstrated that both children and adults use gender agreement to predict nouns when they are preceded by a gender-marked determiner (Brouwer et al., 2017; Cholewa et al., 2019; Lew-Williams & Fernald, 2007). Far less evidence of this has been found in the second language (Hopp & Lemmerth, 2016; Lemmerth & Hopp, 2019). In a primed lexical decision task using auditory homophone primes and orthographic targets, Spinelli and Alario (2002) found that gender-marked determiners constrained lexical access to the gender compatible candidate for French homophones. However, subsequent work provided evidence that grammatical gender does not in fact constrain lexical access but acts at a later stage during the selection of the appropriate candidate (Spinelli et al., 2006). Our results clearly demonstrate that the L1 gender of stored words was activated and interacted with the L2 gender of actually presented words. It remains to be determined whether such was due to the presence of the salient and reliable morphological marking carried in the determiner (i.e. whether participants computed gender based on morpho-phonological cues) or due to the activation of the L2 gender from the auditory noun itself (i.e. retrieval of gender from a newly stored representation). Nevertheless, our results suggest that gender congruency across languages affected lexical access for newly learned L2 nouns, which was hindered when competing gender features from the L1 were activated.

Our results, showing a clear interaction between the established L1 and a newly acquired L2 vocabulary, add to the ample body of experimental evidence showing that the interaction between languages is ubiquitous and found at all levels of processing (see Kroll & Bialystok, 2013, for a review of experimental studies). Indeed, the automatic activation across languages has been demonstrated during both auditory and written processing at the phonological level (Carrasco et al., 2012; Friesenet al., 2020), the lexical level (Dijkstra et al., 2000; Lagrou et al. , 2011; Sunderman & Kroll, 2006) and the syntactic level (Dussias & Sagarra, 2007; Foucart & Frenck-Mestre, 2011). Moreover, parallel activation across languages is found despite distinct writing systems (Thierry & Wu, 2012; Wu & Thierry, 2010). Our results present further evidence that even in monolingual contexts and even for novice learners, the two languages are activated in parallel.

Differences between ERP and behavioral results are likely due to discrepancies in timing and granularity. Whereas the N400 is time-locked to stimuli in such a way as to provide information regarding early aspects of lexical-semantic processing, behavioral measures can encompass both early and late effects during processing. Otherwise stated, by the time low-temporal resolution behavioral data, such as accuracy or reaction time, are measured, other, later, cognitive processes have plausibly occurred (ex. working and episodic memory processes, response selection, mental imagery, c.f. Hauk, 2016). Specifically pertaining to gender congruency effects during a semantic categorization task in the L2, previous ERP results showed a negative (LAN) modulation for cross-language gender incongruent words but no differences in reaction time between conditions (Boutonnet et al., 2012). Similarly, in our study, an N400 effect was seen for Match vs. Mismatch trials but only for cross-language gender congruent words, suggesting that cross-language gender congruency had a direct effect on semantic processing. Once again, this effect was not seen in behavioral results (d prime or accuracy). We posit that in both our and Boutonnet et al.'s (2012) studies, cross-language gender interference effects happen too early in lexical processing to be evidenced behaviorally, which explains why the GCE as measured by EEG (N400) was not echoed by a behavioral effect. Our results are also in line with those reported by McLaughlin and colleagues (2004) who found cortical sensitivity to the lexical status of newly learned L2 items in the absence of behavioral capacity to distinguish between words and non-words. In similar fashion our results show cortical sensitivity to gender congruency across languages that was not reflected in the behavioral response.

Last, we can note that our design did not allow us to examine the effect of gender concord or gender congruency within the L2. Several studies are warranted to delve further into this question. It would be of interest, for example, to test whether we could establish gender congruency effects within the newly learned L2 by adopting a paradigm similar to that used by Boutonnet et al. (2012), in which successive trials carried the same L2 gender, or a visual world paradigm (Hopp & Lemmerth, 2016) in which the gender of the items is varied both across languages and within the L2. Concerning gender concord, a typical violation paradigm could be added to the current design, whereby the learned L2 nouns are preceded by either the correct determiner or the incorrect determiner (Foucart & Frenck-Mestre, 2011, 2012; Morgan-Short et al., 2010). We demonstrated previously that the overlap of grammatical features between the learners' L1 and L2, the specific syntactic structure and the level of proficiency in the L2 all play a role in the pattern of ERP components elicited by gender concord violations in sentential contexts for L2 speakers (Foucart & Frenck-Mestre, 2011, 2012). In similar fashion, recent work that manipulated gender concord in sentential contexts found that the ERP signature for these violations – generally the LAN/P600 – can vary even in native speakers as a function of stimulus characteristics (Beatty- Martínez, et al., 2021). This challenges the notion that ERP signatures associated with syntactic processing (error detection) are tied to a specific level of processing and/or are more systematic in native speakers than in L2 learners. Hence, in novice learners of a gendered language, whether we would find an ERP response to gender concord errors within the newly learned language and, if so, what component would be elicited is an open question.

To conclude, we have provided clear electrophysiological evidence of gender congruency effects in an L2, from the very beginning of acquisition. The clear N400 effect elicited by mismatched compared to matched audio-visual pairs for gender congruent trials was basically annulled for gender incongruent trials. To our knowledge, no prior work has provided evidence of GCE, either behaviorally or via the cortical response to newly learned L2 words. It is important to note that our behavioral results clearly demonstrate that participants learned the correct association between auditory words and images and that this was independent of both the gender in the second language vocabulary and, importantly, of the congruency of gender across the L2 and the participants' L1. Hence, we have also provided evidence of the importance of a multi-disciplinary approach to bring the effects of cross-linguistic gender congruency to light, at least in the early stages of acquisition.

Tables.

Fixed effects: Midline Sites	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.24420	0.53291	2.335	0.0328*
COND.sum1	0.43901	0.20430	2.149	0.0442*
GEND.sum1	-0.19069	0.10448	-1.825	0.0680
EXPE.sum1	0.04772	0.10497	0.455	0.6494
COND.sum1:GEND.sum1	-0.30688	0.13645	-2.249	0.0338*
GEND.sum1:EXPE.sum1	-0.10439	0.10450	-0.999	0.3179
COND.sum1:EXPE.sum1	0.47717	0.10474	4.556	5.31e-06 ***
COND.sum1:GEND.sum1:EXPE.sum1	-0.42132	0.10454	-4.030	5.63e-05 ***
Fixed effects: Frontal Central Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.191e-01	4.801e-01	1.915	0.07348
COND.sum1	3.745e-01	1.859e-01	2.015	0.06091
GEND.sum1	-2.213e-01	6.605e-02	-3.350	0.00255 **
EXPE.sum1	5.209e-02	6.349e-02	0.820	0.41200
COND.sum1:GEND.sum1	-2.581e-01	6.322e-02	-4.082	4.48e-05
GEND.sum1:EXPE.sum1	-1.010e-01	6.325e-02	-1.597	0.11020
COND.sum1:EXPE.sum1	4.016e-01	6.344e-02	6.331	2.50e-10**
COND.sum1:GEND.sum1:EXPE.sum1	-2.057e-01	6.326e-02	-3.251	0.00115 **
Fixed effects: Central Parietal Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.525e-01	4.365e-01	1.266	0.224
COND.sum1	4.020e-01	1.799e-01	2.234	0.040 *
GEND.sum1	-7.605e-02	6.285e-02	-1.210	0.226
EXPE.sum1	-8.894e-03	6.317e-02	-0.141	0.888
COND.sum1:GEND.sum1	-2.463e-01	6.285e-02	-3.919	8.92e-05***
GEND.sum1:EXPE.sum1	-3.943e-02	6.286e-02	-0.627	0.530
COND.sum1:EXPE.sum1	4.686e-01	6.312e-02	7.424	1.19e-13**
COND.sum1:GEND.sum1:EXPE.sum1	-2.634e-01	6.287e-02	-4.189	2.81e-05**

Model: lmer(MVC ~(1+COND.sum|SUBJECTS) + (1+COND.sum|ITEM) + COND.sum + GEND.sum + EXPE.sum + COND.sum:GEND.sum + GEND.sum:EXPE.sum + COND.sum:EXPE.sum + GEND.sum:EXPE.sum:COND.sum)

Table 3a. *Model output for mean voltage ERPs for L2 learners as a function of Session (Pre vs. Post), Gender (Same vs. Opposite) and Condition (Match vs. Mismatch).*

Fixed effects Midline Sites				
	Estimate	Std. Error	t value	Pr(>|t|
(Intercept)	1.20538	0.58093	2.075	0.0533
COND.sum1	-0.02400	0.22892	0.105	0.9177
GEND.sum1	-0.09297	0.17807	0.522	0.6061
COND.sum1:GEND.sum1	0.06651	0.14335	0.464	0.6427
Fixed effects Frontal Central Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.775e-01	5.270e-01	1.665	0.115
COND.sum1	7.535e-03	1.714e-01	0.044	0.965
GEND.sum1	-1.183e-01	8.777e-02	-1.347	0.178
COND.sum1:GEND.sum1	-8.305e-02	8.783e-02	-0.946	0.344
Fixed effects Central Parietal Sites				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.58026	0.44970	1.290	0.215
COND.sum1	-0.05230	0.20715	-0.252	0.804
GEND.sum1	-0.03018	0.08599	-0.351	0.726
COND.sum1:GEND.sum1	-0.01232	0.08601	-0.143	0.886

Model: lmer(MVC ~(1+COND.sum|SUBJECTS) + (1+COND|ITEM) + COND.sum + GEND.sum + COND.sum:GEND.sum, data = d)

Table 3b. *Model output for mean voltage ERPs for L2 learners pre-training as a function of Gender (Same vs. Opposite) and Condition (Match vs. Mismatch).*

Fixed effects: Midline	Estimate	Std. Er	t value	Pr(>|t|)
(Intercept)	0.4832	0.5397	1.117	0.28032
COND.sum1	0.8596	0.2227	3.859	0.00133**
GEND.sum1	-0.1175	0.0911	1.289	0.19730
COND.sum1:GEND.sum1	-0.5198	0.0911	-5.705	0.00000001
Fixed effects:Frontal Central	Estimate	Std. Er	t value	Pr(>|t|)
(Intercept)	0.95444	0.53177	1.795	0.09117
COND.sum1	0.74904	0.24744	3.027	0.00779**
GEND.sum1	-0.3247	0.10393	-3.125	0.00433**
COND.sum1:GEND.sum1	-0.48185	0.09068	-5.314	0.00000013
Fixed effects: Central Parietal	Estimate	Std. Er	t value	Pr(>|t|)
(Intercept)	0.5397	0.4832	1.117	0.28032
COND.sum1	0.8596	0.2227	3.859	0.00133 **
GEND.sum1	-0.1175	0.0911	-1.289	0.19730
COND.sum1:GEND.sum1	-0.5198	0.0911	-5.705	0.00000001

Model: lmer(MVC ~(1+COND.sum|SUBJECTS) + (1+ COND.sum |ITEM) + COND.sum + GEND.sum + COND.sum:GEND.sum)

Table 3c. *Model output for mean voltage ERPs for L2 learners post-training as a function of Gender (Same vs. Opposite) and Condition (Match vs. Mismatch)*

Simple effect of CONDITION for Same Gender trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.2598	0.6633	4.915	0.000146***
CONDmis	-3.4028	0.7703	-4.417	0.000525***
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.559	0.7118	3.596	0.00225
CONDmis	-2.6034	0.6816	-3.820	0.00151**
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.0111	0.5935	3.388	0.003669 **
CONDmis	-2.9107	0.5770	-5.045	0.000127***
Simple effect of CONDITION for Opposite Gender trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.0730	0.8901	1.205	0.244
CONDmis	-0.2687	0.7569	-0.355	0.727
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8607	0.8055	1.069	0.301
CONDmis	-0.5241	0.6441	-0.814	0.427
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.7543	0.7770	0.971	0.346
CONDmis	-0.5932	0.6129	-0.968	0.347

Model: lmer(MVC ~(1+COND|SUBJECTS) + (1+ COND|ITEM) + COND)

Table 3d. *Model output for mean voltage ERPs for L2 learners post-training as a function of Condition (Match vs. Mismatch) for nouns of the opposite (above) and same gender (below) across languages.*

Simple effect of Gender for Match trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.1568	0.7233	1.599	0.126
GENDsame	2.0753	0.4222	4.916	9.62e-07
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8604	0.8033	1.071	0.3000
GENDsame	1.6908	0.5854	2.888	0.0107
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.7621	0.6029	1.264	0.223
GENDsame	1.2745	0.2570	4.960	7.19e-07***
Simple effect of Gender for Mismatch trials				
Fixed effects: Midline				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.8051	0.5736	1.404	0.1754
GENDsame	-0.9082	0.4340	-2.093	0.0365*
Fixed effects:Frontal Central				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.3356	0.5399	0.621	0.543
GENDsame	-0.3838	0.8047	-0.477	0.640
Fixed effects: Central Parietal				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.1590	0.7017	0.227	0.824
GENDsame	1.0637	0.8916	-1.193	0.250

Model: lmer(MVC ~(1+GEND|SUBJECTS) + (1+ GEND|ITEM) + GEND)

Table 3e. *Model output for mean voltage ERPs as a function of Gender (Same vs. Opposite) in the L2 learner group*

Model summary at midline sites				
Fixed effects:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.0468	0.4197	2.494	0.01758
COND.sum1	1.1548	0.2038	5.665	2.59e-06
GEND.sum1	0.1021	0.1255	0.814	0.42275
GROUP.sum1	0.2257	0.4148	0.544	0.59009
COND.sum1:GEND.sum1	0.3167	0.1127	-2.811	0.00626
GEND.sum1:GROUP.sum1	0.1904	0.1081	1.760	0.07838
COND.sum1:GROUP.sum1	0.2434	0.2014	1.209	0.23548
COND.sum1:GEND.sum1:GROUP.sum1	0.4298	0.1081	3.975	7.11e-05
Model summary at frontal central sites				
Fixed effects:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.075e-01	4.017e-01	2.010	0.052510
COND.sum1	9.765e-01	1.872e-01	5.218	9.42e-06***
GEND.sum1	3.382e-01	7.398e-02	4.572	0.000102***
GROUP.sum1	1.450e-01	4.003e-01	0.362	0.719407
COND.sum1:GEND.sum1	2.782e-01	6.611e-02	-4.208	2.80e-05***
GEND.sum1:GROUP.sum1	-1.400e-02	6.585e-02	-0.213	0.831647
COND.sum1:GROUP.sum1	2.248e-01	1.871e-01	1.202	0.237987
COND.sum1:GEND.sumsame:GROUP.sum1	2.039e-01	6.585e-02	3.096	0.001966**
Model summary at central parietal sites				
Fixed effects:	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.025e-01	3.656e-01	0.554	0.58341
COND.sum1	1.237e+00	2.034e-01	6.083	7.26e-07***
GEND.sum1	-1.375e-01	6.475e-02	-2.124	0.03368 *
GROUP.sum1	-3.365e-01	3.656e-01	-0.920	0.36398
COND.sum1:GEND.sum1	-2.873e-01	6.476e-02	-4.436	9.21e-06***
GEND.sum1:GROUP.sum1	-1.978e-02	6.475e-02	0.305	0.76001
COND.sum1:GROUP.sum1	3.761e-01	2.034e-01	1.849	0.07329
COND.sum1:GEND.sum1:GROUP.sum1	2.331e-01	6.476e-02	3.599	0.00032***

Model: lmer(MVC ~(1+COND.sum|PARTICIPANTS) + (1+COND.sum|ITEM) + COND.sum:GEND.sum + GEND.sum:GROUP.sum + COND.sum:GROUP.sum+ GEND.sum:GROUP.sum:COND.sum, data = d)

Table 4a. *Model output for mean voltage ERPs (post-training for L2 learners) as a function of Group (Native vs. L2), Gender (Same vs. Opposite) and Condition (Match vs. Mismatch)*

Model summary: midline sites				
Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.82079	0.60273	1.362	0.190656
COND.sum1	1.39788	0.31593	4.425	0.000387 ***
GEND.sum1	0.08912	0.16977	0.525	0.603694
COND.sum1:GEND.sum1	0.11428	0.15576	0.734	0.463721
Model summary: frontal central sites				
Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.66200	0.59535	1.112	0.281627
COND.sum1	1.20152	0.27855	4.313	0.000475 ***
GEND.sum1	-0.35212	0.09536	3.693	0.000223 ***
COND.sum1:GEND.sum1	-0.07351	0.09536	0.771	0.440817
Model summary: central parietal sites				
Fixed effects:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.13409	0.54457	-0.246	0.808454
COND.sum1	1.61456	0.33473	4.823	0.000157***
GEND.sum1	-0.15744	0.09181	-1.715	0.086405
COND.sum1:GEND.sum1	0.05374	0.09181	0.585	0.558323

Model: lmer(MVC ~(1+COND.sum|PARTICIPANTS) + (1+COND.sum|ITEM) + COND.sum:GEND.sum + GEND.sum:COND.sum, data = d)

Table 4b. . *Model output for mean voltage ERPS for the native control group as a function of Gender (Same vs. Opposite) and Condition (Match vs. Mismatch)*

References

Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531–552.

Alencar de Resende N.C., Mota M.B., Seuren P. (2019) The Processing of Grammatical Gender Agreement in Brazilian Portuguese: ERP Evidence in Favor of a Single Route. *J Psycholinguist Res*. *48*(1),181-198.

Altenberg, E.P. (2005). The perception of word boundaries in a second language. *Second Language Research*, *21*, 325-358.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305.

Barber, H., & Carreiras, M. (2003). Integrating Gender and Number Information in Spanish Word Pairs: An Erp Study. *Cortex*, *39*(3), 465–482.

Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. Journal of Cognitive Neuroscience, *17*, 137-155.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Beatty-Martínez, A. L., Bruni, M. R., Bajo, M. T., & Dussias, P. E. (2021). Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers. *Psychophysiology*, *58*(30).

Bigi, B. (2015). SPPAS - Multi-lingual approaches to the automatic annotation of speech. *The Pho*netician, *111– 112*(ISSN:0741-6164), 54–69.

Boersma, P. & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/

Bordag, D. (2004). Interaction of L1 and L2 systems at the level of grammatical encoding: Evidence from picture naming. *EUROSLA Yearbook*, *4*(1), 203–230.

Bordag, D., & Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism: Language and Cognition, 10*, 299–314.

Bordag, D., & Pechmann, T. (2008). Grammatical gender in translation. *Second Language Research, 24*, 139–166.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a Neurobiologically Plausible Model of Language-Related, Negative Event-Related Potentials. *Frontiers in Psychology*, *10*, 298.

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79

Breitenstein, C., Zwitserlood, P., de Vries, M., Feldhues, C., Knecht, S., & Dobel, C. (2007). Five days versus a lifetime: Intense associative vocabulary training generates lexically integrated words. *Restorative Neurology and Neuroscience*, *25*(5,6), 493–500.

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition*, *41*(2), 281–296.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science, 41*(Suppl 6), 1318–1352.

Brouwer, S., Sprenger, S., & Unsworth, S. (2017). Processing grammatical gender in Dutch: evidence from eye movements *Journal of Experimental Child Psychology*, *159*, 50–65.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177-208.

Carrasco, H., Midgley, K. & Frenck-Mestre, C. (2012). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones. *Psychophysiology*, *49*, 531-543.

Cholewa, J., Neitzle, I., Bürsgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: An eye-tracking study with monolingual German 3rd and 4th graders. *Frontiers in Psychology*, *10*, Article 2586.

Chun, E., Choi, S., & Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study. *Neuroscience Letters*, *521*(2), 125–129.

Clark, E.V. (2009). *First Language Acquisition*. 2nd Edition. Cambridge, UK: Cambridge University Press.

Corrêa L, Augusto M. & Castro A., (2011). Agreement and markedness in the ascription of gender to novel animate nouns by children acquiring Portuguese. *Journal of Portuguese Linguistics*, *10*(1), 121-142.

Corréa, L.M.S., & Name, M.C. (2003). The processing of determiner-noun agreement and the identification of the gender of nouns in the early acquisition of Portuguese. *Journal of Portuguese Linguistics*, *2*, 19–43.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1283–1296.

Costa, A., Kovacic, D., Franck, J., & Caramazza, A. (2003). On the autonomy of the grammatical gender systems of the two languages of a bilingual. *Bilingualism: Language and Cognition*, *6*(3), 181–200.

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M. & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language*, *53*, 42–59.

Dahan, D., Swingley, D., Tanenhaus, M., & Magnuson, J. (2000). Linguistic Gender and Spoken-Word Recognition in French. *Journal of Memory and Language*, *42*, 465–480.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition, 135,* Article 103569. doi.org/10.1016/j.bandc.2019.05.007

Denhovska, N., & Serratrice, L. (2017). Incidental Learning of Gender Agreement in L2. *Journal of Psycholinguistic Research*, *46*(5), 1187–1211.

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On Being Blinded by Your Other Language: Effects of Task Demands on Interlingual Homograph Recognition. *Journal of Memory and Language, 42*(4), 445-464.

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: On the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*(6), 1251–1261.

Dowens, M. G., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish—Evidence from Event Related Potentials. *Neuropsychologia*, *49*(7), 1651–1659.

Dussias, P.E, & Sagarra N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *10*, 101–116.

Foucart, A., Branigan, H. P., & Bard, E. G. (2010). Determiner selection in romance languages: Evidence from French. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1414–1421.

Foucart, A., & Frenck−Mestre, C. (2012). Can late learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*, 226-248.

Foucart, A., & Frenck−Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1 -L2 syntactic similarity. *Bilingualism, Language and Cognition*, *14*, 379-399.

Fraga, I., Padrón & Hinojosa, J. A. (2021) Negative valence effects on the processing of agreement dependencies are mediated by ERP individual differences in morphosyntactic processing, *Language, Cognition and Neuroscience*

Enriquez-Guppert, S., Komrad, C., Pantev, C., & Huster, R.J. (2010) Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *NeuroImage*, *51*, 877-887.

Frenck−Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research*, *21*, 175-198.

Friesen, D. C., Ward, O., Bohnet, J., Cormier, P., & Jared, D. (2020). Early activation of cross-language meaning from phonology during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1754–1767.

Gollan, T., & Frost, R. (2001). Two Routes to Grammatical Gender: Evidence from Hebrew. *Journal of Psycholinguistic Research*, *30*, 627-651

Gunter, T., Friederici, A., & Schriefers, H. (2000). Syntactic Gender and Semantic Expectancy: ERPs Reveal Early Autonomy and Late Interaction. *Journal of Cognitive Neuroscience*, *12*, 556-568.

Hagoort, P. (2003). Interplay between Syntax and Semantics during Sentence Comprehension: ERP Effects of Combining Syntactic and Semantic Violations. *Journal of Cognitive Neuroscience*, *15*(6), 883–899.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61.

Hopp, H., & Lemmerth, N. (2016). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*, 1-29.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*.

Karmiloff-Smith, A. (1979). *A functional approach to child language: A study of determiners and reference*. Cambridge, UK: Cambridge University Press.

Klassen, R. (2016). The representation of asymmetric grammatical gender systems in the bilingual mental lexicon. *Probus*, *28*, 9–28.

Kroll, J.F. & Bialystok, E. (2013). Understanding the Consequences of Bilingualism for Language Processing and Cognition. Journal of Cognitive Psychology (Hove), 25(5):

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205–225.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences:Brain potentials reflect semantic incongruity. *Science, 207,* 203-205.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(1), 1–26.

Lagrou E., Hartsuiker R.J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 952–965.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*(3), 312–330.

Lemmerth, N., & Hopp, H. (2019). Gender processing in simultaneous and successive bilingual children: Cross linguistic lexical and syntactic influences. *Language Acquisition*, *26*(1), 21–45.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38.

Lew-Williams, C., & Fernald, A. (2007). Young Children Learning Spanish Make Rapid Use of Grammatical Gender in Spoken Word Recognition. *Psychological Science*, *18*(3), 193–198.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704.

Meltzer, J., & Braun, A. (2013). P600-like positivity and Left Anterior Negativity responses are elicited by semantic reversibility in nonanomalous sentences. *Journal of Neurolinguistics*, *26*, 129–148.

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement: a reply to Tanner et al. *Cortex*, *66*, 156–159.

Morales, L., Paolieri, D., & Bajo, T. (2011). Grammatical Gender Inhibition in Bilinguals. *Frontiers in Psychology*, *2*. doi.org/10.3389/fpsyg.2011.00284

Morales, L., Paolieri, D., Dussias, P. E., Valdés Kroff, J. R., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism*, *19*(2), 294–310.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, *60*(1), 154–193.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524.

Nichols, E. S., & Joanisse, M. F. (2016). Functional activity and white matter microstructure reveal the independent effects of age of acquisition and proficiency on second-language learning. *NeuroImage*, *143*, 15–25.

Öttl, A., & Behne, D. M. (2017). Assessing the Formation of Experience-Based Gender Expectations in an Implicit Learning Scenario. *Frontiers in Psychology*, *8*. doi.org/10.3389/fpsyg.2017.01485

Paolieri, D., Cubelli R., Macizo, P., Bajo, T., Lotto, L., & Job, R. (2010). Grammatical gender processing in Italian and Spanish bilinguals. *Quarterly Journal of Experimental Psychology, 63*, 1631-1645.

Paolieri, D., Demestre, J., Guasch, M., Bajo, T. & Ferré, P. (2020) The gender congruency effect in Catalan-Spanish bilinguals: Behavioral and electrophysiological evidence. *Biingualism, Language and Cognition*, 23, 1045-1055.

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language, 18*, 571–590.

Popov, S. & Bastiaanse, R. (2018). Processes underpinning gender and number disagreement in Dutch: An ERP study. *Journal of Neurolinguistics*, *46*, 109-121.

Popov, S., Miceli, G., Ćurčić-Blake, B., & Bastiaanse, R. (2020). The role of semantics and repair processes in article noun gender disagreement in Italian: An ERP study. *Brain and Language*, *206*, 104787.

Rodriguez-Fornells, A., & Münte, T.F. (2016). Syntactic interference in bilingual naming during language switching: An electrophysiological study. In J.W. Schweiter (Ed.) *Cognitive Control and Consequences of Multilingualism*, John Benjamins, Amsterdam/Philidelphia. pp. 239-270.

Sá-Leite, A. R., Fraga, I., & Comesaña, M. (2019). Grammatical gender processing in bilinguals: An analytic review. *Psychonomic Bulletin & Review, 26*(4), 1148–1173.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are L1 and L2 processed similarly. *Second Language Research*, *24*, 397–430.

Sabourin, L., Stowe, L.A., & de Haan, G.J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*, 1–29

Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, *10*(3), 257–275.

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., & Martín-Loeches, M. (2014). Differential Task Effects on N400 and P600 Elicited by Semantic and Syntactic Violations. *PLoS ONE*, *9*(3), e91226.

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*(1), 169–194.

Seigneuric, A., Zagar, D., Meunier, F. & Spinelli, S. (2007). The relation between language and cognition in children aged 3 to 9 : The acquisition of grammatical gender in French. *Journal of Experimental Child Psychology*, 96, 229-246.

Soskey, L., Holcomb, P. J., & Midgley, K. J. (2016). Language effects in second-language learners: A longitudinal electrophysiological study of Spanish classroom learning. *Brain Research*, *1646*, 44–52.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Spinelli, E., & Alario, F.-X. (2002). Gender Context Effects on Homophone Words. *Language and Cognitive Processes*, 457–469.

Spinelli, E., Meunier, F., & Seigneuric, A. (2006). Spoken word recognition with gender-marked context. *The Mental Lexicon*, 277–297.

Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *NeuroImage*, *33*(2), 774–783.

Sunderman G., & Kroll, J.F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, 28, 387–422.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on "Grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex*, *66*, 149–155.

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367–382.

Thierry, G., & Wu, Y. (2017). Brain potentials reveal unconscious translation during foreign-language comprehension. Proceedings of the National Academy of Science, *104*: 12530±12535

Tokowicz, N., & MacWhinney, B. (2005). Implicit and Explicit Measures of Sensitivity to Violations in Second Language Grammar: An Event-Related Potential Investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204.

Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288.

Wu., X & Thierry, G. (2010). Chinese–English Bilinguals Reading English Hear Chinese (2010). *The Journal of Neuroscience*, 30, 7646 –7651

Yum, Y., Midgley, K., Holcomb, P., & Grainger, J. (2014). An ERP study on initial second language vocabulary learning. *Psychophysiology*, 51(4), 364–373.

Appendix: Nouns that were learned in the training sessions. The nouns included in the pre- and post-training Match/Mismatch task are indicated with an asterisk.

BR Portuguese	French	English
	Same Gender (Portuguese/French)	
*janela (f)	fenêtre(f)	window
*lareira (f)	cheminée(f)	fireplace
*casaco (m)	manteau(m)	coat
*saia (f)	jupe(f)	skirt
*isqueiro	briquet(m)	lighter
*chaleira(f)	bouilloire(f)	kettle
*panela(f)	poële(f)	pan
*linguiça(f)	saucisse(f)	sausage
*chão(m)	sol(m)	floor
*estante(f)	étagère(f)	bookcase
*peneira(f)	passoire(f)	colander
*pau(m)	bâton(m)	stick
*gancho(m)	crochet(m)	hook
* escova(f)	brosse(f)	brush
*trapo(m)	chiffon(m)	rag
barata(f)	blatte(f)	cockroach
frango(m)	poulet(m)	chicken (food)
capacho(m)	tapis(m)	rug
	Opposite Gender (Portuguese/French)	
*vagem(f)	haricot vert(m)	greenbean
*alface(m)	laitue(f)	lettuce
*brinco(m)	boucle d'oreille(f)	earring
*garfo(m)	fourchette(f)	fork
*vassoura (f)	balai(m)	broom
*calça(f)	pantalon(m)	pants
*tigela(f)	bol(m)	bowl
*caneta(f)	stylo(m)	pen
*lixo(m)	poubelle(f)	tashcan
*giz(m)	craie(f)	chalk
*faca(f)	couteau(m)	knife
*luva(f)	manique(f)	oven mitt
*gaveta(f)	tiroir(m)	drawer
*parafuso(m)	tournevis(m)	screw
*cachimbo(m)	pipe(f)	pipe
pasta(f)	classeur(m)	folder
camundongo(m)	souris(f)	mouse
guardanapo(m)	serviette(f)	napkin